

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
14 June 2001 (14.06.2001)

PCT

(10) International Publication Number
WO 01/42432 A2(51) International Patent Classification⁷: C12N 9/00 (74) Agents: WALTON, Séan, M. et al.; Mewburn Ellis, York House, 23 Kingsway, London, Greater London WC2B 6HP (GB).

(21) International Application Number: PCT/GB00/04661

(22) International Filing Date: 6 December 2000 (06.12.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
9929061.1 8 December 1999 (08.12.1999) GB

(71) Applicant (for all designated States except US): MEDICAL RESEARCH COUNCIL [GB/GB]; 20 Park Crescent, London, Greater London W1N 4AL (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): FERSHT, Alan [GB/GB]; Department of Chemistry, Cambridge University, Lensfield Road, Cambridge, Cambridgeshire CB2 1EW (GB). ALTAMIRANO, Myriam [MX/GB]; Centre for Protein Engineering, MRC Centre, Hills Road, Cambridge, Cambridgeshire CB2 2QH (GB).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 01/42432 A2

(54) Title: METHODS OF PRODUCING NOVEL ENZYMES

(57) Abstract: Methods of obtaining enzymes that bind target substrate and catalyse desired reactions. α/β -barrel proteins are categorised into two classes based on catalytic lid structure. Lids can be grafted onto scaffolds with additional minor modifications at conserved and non-conserved residues to provide candidate product enzymes for screening for the desired properties. Design of a novel enzyme which binds a target substrate and catalyses a reaction of choice is facilitated by selection of a scaffold which binds the substrate and of a catalytic lid of the correct class for the desired reaction. Targeted or focussed mutagenesis may be used to refine substrate binding and catalysis.

METHODS OF PRODUCING NOVEL ENZYMES

The present invention relates to protein design, specifically design of enzymes. It is based on work of the inventors in categorising α/β -barrel proteins into two classes based on catalytic lid structure, and recognising that enzymes which catalyse a given class of reactions are found in one or other of the two classes. Design of a novel enzyme which binds a target substrate and catalyses a reaction of choice is facilitated by selection of a scaffold which binds the substrate and of a catalytic lid of the correct class for the desired reaction. Targeted or focussed mutagenesis may be used to refine substrate binding and catalysis.

Enzymes are Nature's catalysts. They are proteins that have evolved to bind specific substrates and catalyse specific reactions at optimal efficiency and yield under conditions in the cell. However, using protein engineering only a few highly active new enzymes have been produced, and no general methodology achieved. Such catalysts as have been made have employed specific features unique to individual proteins (*Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. A. Fersht (WH Freeman and Co, 1999), chapters 15 and 16). The field of catalytic antibodies in which the naturally binding proteins have been evolved to become catalysts has also failed in general to produce highly active molecules that rival natural enzymes (Fersht, *supra*, pp 60, 361).

Natural evolution involves mutation and selection. Random mutation and selection *in vitro* is, without simplifying rules, too difficult and time consuming because a large

number of mutations have generally to be made to evolve a new catalytic activity. The present inventors have appreciated that Nature has evolved design principles to diversify α/β -barrel protein activity more rapidly, and here 5 provide rules for novel enzyme design that greatly reduce the number and choice of residues which to mutate.

The α/β -barrel

Proteins adopt many different topologies of folded 10 structure. However, one particular fold, the α/β -barrel, is the most common, accounting for some 10% of known enzymes. The α/β -barrel is clearly an important target as the framework for novel protein design, but despite considerable efforts no one has deciphered and demonstrated 15 experimentally how Nature is able to use this design of fold so effectively.

It has been speculated previously that the binding sites in 20 α/β -barrel enzymes may have evolved by divergent evolution, so acquiring the ability to bind other substrates (cited in Fersht, *supra*). Specifically, an archetypal enzyme that catalyses a particular reaction on a particular substrate may evolve into a family of enzymes catalysing the same reaction, but on a variety of substrates.

25 The inventors have analysed a particular structure of α/β -barrel enzymes, called the "active-site lid", that is involved primarily with catalysis rather than specificity of binding (see below). The lid contains amino acid residues 30 whose function is providing catalytic chemical groups in the active site. The lids are herein divided into two main classes. The inventors have identified a correlation

between the class of the lid and the kind of mechanism catalysed by the enzyme. From this, the present invention provides for grafting a template lid onto a selected barrel framework, or modifying an underlying framework to provide 5 an altered lid (e.g. a lid of the alternative class), and then subjecting the lid to targeted mutagenesis and selection, to create new enzymes catalysing a desired reaction.

10 **GLOSSARY**

Helix

A helix is formed by a polypeptide chain with repeating phi and psi angles. Its geometry is defined by the number of residues per turn, and the rise per residue. In principle 15 the polypeptide chain can form right and left handed helices with a range of pitches (see Fersht, *supra*, and *Introduction to Protein Structure*, 2nd. Edition Bränden, C., and Tooze, J. (Garland Publishing Inc., New York, 1999)).

20 *Loop*

A protein loop is any stretch of nonregular polypeptide chain connecting secondary structures. Short loop regions adopt a restricted set of conformations and loop families have been recognised in specific supersecondary structures.

25

Beta Sheet (β sheet)

These structures are formed from residues in an extended conformation with psi phi bond angle pairs in the wide allowed region in the upper left hand corner of the 30 Ramachandran plot. The strands of the beta sheet are not fully extended, due to the constraints of hydrogen bonding, and the sheets appear pleated. In addition there is a left-handed twist between adjacent strands when looking at right

angles to the strand direction (Chothia, 1973, *J. Mol. Biol.* 75: 295-302). The beta strands in a sheet can be arranged to form parallel, antiparallel or mixed sheets. Refer to Richardson, (1977) *Nature* 268: 495-500.

5

Beta Strand (β Strand)

A beta strand describes a single length of polypeptide chain that forms part of a beta sheet.

10 *Parallel Beta Sheets*

This is a beta-pleated sheet in which successive beta strands all lie parallel in three dimensions. Such sheets have evenly spaced hydrogen bond pairs that lie at an angle to the beta strands

15

Beta-Alpha-Beta Units:

Beta-alpha-beta units consist of two parallel hydrogen bonded beta strands connected by a loop containing at least one alpha helix.

20

Beta Barrel

In some instances large anti-parallel (or parallel) sheets can roll up completely to join edges and form a cylinder or closed 'barrel', in which the first strand is hydrogen bonded to the last.

25

Topology (fold family),

Structures are grouped into fold families at this level depending on both the overall shape and connectivity of the 30 secondary structures. This is done using the structure comparison algorithm SSAP (Taylor and Orengo (1989) *J. Mol. Biol.* 208: 1-22 and (1989) *Protein Eng.* 2: 505-519.

Parameters for clustering domains into the same fold family have been determined by empirical trials throughout the Brookhaven databank. Structures which have a SSAP score of 70 and where at least 60% of the larger protein matches the 5 smaller protein are assigned to the same T level or fold family.

Topology cartoons

10 Protein topology cartoons are simplified representations of protein folds. These diagrams are two-dimensional schematic representations of protein structures. They represent the structure as a sequence of secondary structure elements (helices and strands), and illustrate the relative spatial position and direction of these elements.

15

Homologous Superfamily

20 This level groups together protein domains which are thought to share a common ancestor and can therefore be described as homologous. Similarities are identified first by sequence comparisons and subsequently by structure comparison using SSAP. Structures are clustered into the same homologous superfamily if they satisfy one of the following criteria.

- 25 (i) Sequence identity $\geq 35\%$, 60% of larger structure equivalent to smaller
- (ii) SSAP score ≥ 80.0 and sequence identity $\geq 20\%$
- (iii) 60% of larger structure equivalent to smaller
- (iv) SSAP score ≥ 80.0 , 60% of larger structure equivalent to smaller, and
- 30 (iv) domains have related functions

Sequence Families

Structures within each homologous superfamily are further

clustered on sequence identity, using CATH (see below). Domains clustered in the same sequence families have sequence identities >35% (with at least 60% of the larger domain equivalent to the smaller), indicating highly similar structures and functions. (Thornton et al., *J. Mol. Biol.* 293, 333-342. (1999); Taylor and Orengo *J.Mol. Biol.* 208. 1-22. (1989a); Taylor and Orengo *Protein Eng.* 2. p.505-519. (1989b)).

10 *Active site lid in α/β -barrel proteins*

This is the structure that covers the active site, closing and shielding it from solvent.

α/β -barrel proteins are identified in the CATH and SCOP databases (CATH - A Hierarchic Classification of Protein Domain Structures, Orengo et al. *Structure*. 5, 1093-1108 (1997) <http://www.biochem.ucl.ac.uk/bsm/cath/> ; SCOP - Murzin et al., *J. Mol. Biol* 247:536-540 (1995) and see also <http://scop.mrc-lmb.cam.ac.uk/scop>), and in the dedicated 20 database for such proteins TIM-DB at <http://argo.urv.es/~pujadas/TIM/> and Pujadas & Palau *Biologia, Bratislava*, 54 (3): 231-254, (1999)..

25 A list of α/β -barrel proteins to which aspects of the present invention can be applied, or which can be employed in the present invention, appears below as Table IV. Each of these has a scaffold including a binding site for a substrate or ligand, and an active site lid. In accordance with the present invention the scaffold or binding site of 30 any of these may be employed either to bind a substrate of choice or as a starting point for mutagenesis and selection for ability to bind the chosen substrate. Likewise, the active site lid of any of these may be grafted onto a chosen

scaffold and employed either to catalyse the desired reaction on the chosen substrate or as a starting point for mutagenesis and selection for ability to catalyse the desired reaction. As explained, an active site lid for a 5 desired reaction or type of reaction may be chosen at least partly on the basis of its classification as a Class I or Class II α/β -barrel as defined herein.

Table III shows an overview of different reaction mechanisms 10 for which α/β -barrel enzymes have been found to be active. In selection of a particular architecture for the active site in accordance with the present invention, the kind of reaction mechanism involved (e.g. proton abstraction, protein abstraction after enolisation, proton abstraction 15 from Schiff base intermediates, metal activated hydrolysis, attack of amino-acid side-chain nucleophiles to specifically activated atoms in the substrate, and so on) may be taken into account. Thus, where a reaction of a particular type is desired, an active site lid of the 20 appropriate class may be selected, preferably an active site lid which catalyses the desired reaction or a similar reaction (albeit with a different substrate).

All documentation cited herein is incorporated by reference, 25 including internet sites and databases (especially in the form available at the date of filing of the present specification, but where possible including the latest updates).

Brief Description of the Figures

30 Figure 1 shows schematic representation and structural features of the two classes of α/β barrel proteins, illustrated with reference to PRAI (Class I) and IGPS (Class II). The eight β -strands of the barrel are indicated by

triangles. Alpha helices are indicated by rectangles and the constant regions, phosphate binding ($\beta 7\alpha 7$ and $\beta 8\alpha 8$) and the anthranilate binding site ($\beta 2\alpha 2$), by dark loops. For Class I (PRAI group) structures, the main feature of the

5 active site lid ($\beta 6\alpha 6$) is represented by the loop in white with a shadow. The structure is a view from the top of the barrel which constitutes the active site of PRAI. The lotus leaf lid $\beta 6\alpha 6$ is indicated by a white ribbon. The $\beta 1\alpha 1$ loop is the shorter of the two white ribbons. The constant

10 regions (phosphate binding site and anthranilate binding site) are shaded. The clover leaf (shadow) lid of the Class II structure is also shown, which has three principal elements: the extra N-terminal; loop $\beta 1\alpha 1$; and loop $\beta 6\alpha 6$ (all dark). The other structural features are indicated as above.

15 The structure is a top view of the Class II (IGPS group) barrel. The IGPS scaffold, extra N-terminal residues, and the $\beta 1\alpha 1$ and $\beta 6\alpha 6$ loop are indicated by dark ribbons. The constant regions are shaded.

20 Figure 2 illustrates the reactions catalysed by phosphoribosyl anthranilate isomerase (PRAI) and indoleglycerol-phosphate synthase (IGPS). The PRAI reaction is an intramolecular redox reaction (Amadori rearrangement) of N-5-phosphoribosyl)anthranilate (PRA) to (1-(2-

25 carboxyphenylamino)-1-deoxyribulose 5-phosphate (CdRP). In the IGPS reaction, the substrate CdRP undergoes an irreversible ring-closure to indoleglycerol phosphate (IGP) with release of CO_2 and H_2O . Chemical reduction of CdRP by borohydride produces the substrate analogue rCdRP for IGPS.

30 The rCdRP is an inhibitor of both enzymes.

Figure 3 shows a sequence alignment of *in vitro* evolved PRAI (*ivePRAI*), PRAI and IGPS. The single-letter code for amino acid residues is used. Residues in IGPS (71-254) Identities 167/184 (90%); similarities: 171/184 (92%). Residues in PRAI (375-396) 8/18 (44%); similarities 12/18 (66%). Identities: 5 outline, bold and shade; Similarities: outline and shade

Figure 4 shows a protein topology ("TOPS") cartoon for a protein (triangular symbols represent beta strands and the 10 circular ones helices).

Figure 5 shows a protein topology ("TOPS") cartoon for another protein (triangular symbols represent beta strands and the circular ones helices).

15 Figure 6 illustrates topology of a protein with reference to its sequence.

Sequence identities are calculated using the program Blast, 20 using the following parameters: H=0, V=-20, B= 20, S=40, - ctxfactor=1.00, E=64.8038 (Altschul et al., (1990) *J. Mol. Biol.* 215: 403-410).

Aspects and embodiments of the present invention are 25 disclosed throughout this text, and generally provide methods of obtaining novel enzymes, or in particular methods of obtaining an enzyme that catalyses a desired reaction on a target substrate. The invention also provides a method of classifying α/β -barrel proteins into two classes by means of 30 applying criteria disclosed herein, and a method whereby an α/β -barrel protein is appointed as a member of Class I or Class II in accordance with these criteria. Following classification, a method according to the invention may

generally provide for alteration of the active site lid of an α/β -barrel protein of Class I to convert it into Class II, or may generally provide for alteration of the active site lid of an α/β -barrel protein of Class II to convert it into Class I. Moreover, the present invention provides for modification of an α/β -barrel protein which catalyses a first reaction of a given reaction type into an α/β -barrel protein which catalyses a second reaction of that reaction type, and also provides for modification of an α/β -barrel protein which catalyses a first reaction of a given reaction type into an α/β -barrel protein which catalyses a second reaction of a different reaction type. By means of one or more of such methods, an enzyme which catalyses a desired reaction on a target substrate may be obtained, and this may involve conversion of an enzyme from one of Class I and Class II to the other (especially where a protein is modified to catalyse a reaction of a different type), or may involve maintenance of a structure conforming to Class I or to Class II, while altering substrate binding specificity and/or reaction catalysed.

A method of obtaining an enzyme in accordance with the present invention may involve modifying one or more, or preferably a combination of the following regions: the N-terminal segment, the β_1 - α_1 loop, and the β_6 - α_6 loop, especially where an enzyme of one of Class I and Class II is converted into the other Class. In preferred embodiments, one or more of the following may additionally be mutated: extra domains between β_3 α_3 and C-terminal segment (after β_8).

30

As is discussed in detail elsewhere herein, a scaffold may be chosen (for engineering of a desired active site lid)

from any α/β -barrel protein, but is preferably chosen to be one which binds the target substrate of interest. Where such a scaffold is not available, a second preference is for a α/β -barrel protein which binds a similar substrate, i.e. a 5 molecule with as much structural similarity as possible. Mutation of the scaffold may then be used to alter its binding specificity so it binds the target substrate. The regions which may be mutated in order to alter substrate binding specificity are discussed elsewhere herein.

10

A method of obtaining an enzyme in accordance with the present invention may be used to provide a protein which comprises an α/β -barrel scaffold which binds a target substrate and a catalytic lid which catalyses a desired 15 reaction. The scaffold may be provided from a α/β -barrel which naturally binds said target substrate, or may be provided by a method comprising mutation of a α/β -barrel and selection for binding to said target substrate. Such enzymes are provided as further aspects of the present 20 invention, as is their use in a method of catalysing the desired reaction on the target substrate, along with other aspects and embodiments disclosed herein.

A protein or polypeptide according to the present invention 25 may be considered "chimaeric", in embodiments where the scaffold is of one protein and the active site lid is of another protein. The resultant chimaera may represent a "humanised" enzyme, wherein a human enzyme is modified to introduce an enzymatic activity of a non-human, e.g. other 30 mammalian or microbial, enzyme. The present invention allows for minimal, minor modification to a parent scaffold (e.g. human) to introduce the desired enzymatic activity,

minimising effects on immunogenicity in a human of the product enzyme. Usually, in addition to grafting of an active site lid onto a scaffold, or engineering a protein with a particular scaffold to alter its active site lid, 5 some further mutation may be required to obtain the desired catalysis on the target substrate or may be desirable to increase affinity for substrate and/or rate of catalysis. Appropriate regions of proteins for such targeted mutation are discussed in detail elsewhere herein, and include 10 catalytic residues, β 1- α 1 loop and/or β 2- α 2 loop (for Class I), metal binding site, N-terminal extension and/or C-terminal extension (for Class II).

A suitable selection system may be employed to identify 15 mutations with the desired effect. For instance, phage display may be used to identify members of a population of mutated proteins which bind a target substrate. Selection systems, including *in vivo* selection systems, for catalysis of the desired reaction may be available or can be designed, 20 as exemplified experimentally below.

A convenient way of producing a polypeptide according to the present invention is to express nucleic acid encoding it, by use of nucleic acid in an expression system.

25 Accordingly the present invention also provides in various aspects nucleic acid encoding the polypeptides of the invention, which may be used for production of the encoded polypeptide.

30 Generally when encoding for a polypeptide in accordance with the present invention, nucleic acid is provided as an isolate, in isolated and/or purified form, or free or substantially free of material with which it is naturally

associated, such as free or substantially free of nucleic acid flanking the gene in the human genome, except possibly one or more regulatory sequence(s) for expression. Nucleic acid may be wholly or partially synthetic and may include 5 genomic DNA, cDNA or RNA. Where nucleic acid according to the invention includes RNA, reference to the sequence shown should be construed as encompassing reference to the RNA equivalent, with U substituted for T.

10 Nucleic acid sequences encoding a polypeptide in accordance with the present invention can be readily prepared by the skilled person using the information and references contained herein and techniques known in the art (for example, see Sambrook, Fritsch and Maniatis, "Molecular 15 Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press (1989), and Ausubel *et al.*, Current Protocols in Molecular Biology, John Wiley and Sons, (1994)), given the nucleic acid sequence and clones available. These techniques include (i) the use of the polymerase chain 20 reaction (PCR) to amplify samples of such nucleic acid, e.g. from genomic sources, (ii) chemical synthesis, or (iii) preparing cDNA sequences. DNA encoding a polypeptide may be generated and used in any suitable way known to those of skill in the art, including by taking encoding DNA, 25 identifying suitable restriction enzyme recognition sites either side of the portion to be expressed, and cutting out said portion from the DNA. The portion may then be operably linked to a suitable promoter in a standard commercially available expression system. Another recombinant approach 30 is to amplify the relevant portion of the DNA with suitable PCR primers. Modifications to the relevant sequence may be made, e.g. using site directed mutagenesis, to lead to the expression of modified polypeptide or to take account of

codon preference in the host cells used to express the nucleic acid.

In order to obtain expression of the nucleic acid sequences,
5 the sequences may be incorporated in a vector having one or more control sequences operably linked to the nucleic acid to control its expression. The vectors may include other sequences such as promoters or enhancers to drive the expression of the inserted nucleic acid, nucleic acid
10 sequences so that the polypeptide is produced as a fusion and/or nucleic acid encoding secretion signals so that the polypeptide produced in the host cell is secreted from the cell. Polypeptide can then be obtained by transforming the vectors into host cells in which the vector is functional,
15 culturing the host cells so that the polypeptide is produced and recovering the polypeptide from the host cells or the surrounding medium. Prokaryotic and eukaryotic cells are used for this purpose in the art, including strains of *E. coli*, yeast, and eukaryotic cells such as COS or CHO cells.

20 Thus, the present invention also encompasses a method of making a polypeptide (as disclosed), the method including expression from nucleic acid encoding the polypeptide (generally nucleic acid according to the invention). This
25 may conveniently be achieved by growing a host cell in culture, containing such a vector, under appropriate conditions which cause or allow expression of the polypeptide. Polypeptides may also be expressed in *in vitro* systems, such as reticulocyte lysate.

30 Systems for cloning and expression of a polypeptide in a variety of different host cells are well known. Suitable host cells include bacteria, eukaryotic cells such as

mammalian and yeast, and baculovirus systems. Mammalian cell lines available in the art for expression of a heterologous polypeptide include Chinese hamster ovary cells, HeLa cells, baby hamster kidney cells, COS cells and many others. A common, preferred bacterial host is *E. coli*.

Suitable vectors can be chosen or constructed, containing appropriate regulatory sequences, including promoter sequences, terminator fragments, polyadenylation sequences, 10 enhancer sequences, marker genes and other sequences as appropriate. Vectors may be plasmids, viral e.g. 'phage, or phagemid, as appropriate. For further details see, for example, Molecular Cloning: a Laboratory Manual: 2nd edition, Sambrook et al., 1989, Cold Spring Harbor 15 Laboratory Press. Many known techniques and protocols for manipulation of nucleic acid, for example in preparation of nucleic acid constructs, mutagenesis, sequencing, introduction of DNA into cells and gene expression, and analysis of proteins, are described in detail in Current 20 Protocols in Molecular Biology, Ausubel et al. eds., John Wiley & Sons, 1992.

Thus, a further aspect of the present invention provides a host cell containing encoding nucleic acid as disclosed 25 herein.

The nucleic acid of the invention may be integrated into the genome (e.g. chromosome) of the host cell. Integration may be promoted by inclusion of sequences which promote 30 recombination with the genome, in accordance with standard techniques. The nucleic acid may be on an extra-chromosomal vector within the cell, or otherwise identifiably heterologous or foreign to the cell.

A still further aspect provides a method which includes introducing the nucleic acid into a host cell. The introduction, which may (particularly for *in vitro* introduction) be generally referred to without limitation as "transformation", may employ any available technique. For eukaryotic cells, suitable techniques may include calcium phosphate transfection, DEAE-Dextran, electroporation, liposome-mediated transfection and transduction using retrovirus or other virus, e.g. vaccinia or, for insect cells, baculovirus. For bacterial cells, suitable techniques may include calcium chloride transformation, electroporation and transfection using bacteriophage. As an alternative, direct injection of the nucleic acid could be employed.

Marker genes such as antibiotic resistance or sensitivity genes may be used in identifying clones containing nucleic acid of interest, as is well known in the art.

The introduction may be followed by causing or allowing expression from the nucleic acid, e.g. by culturing host cells (which may include cells actually transformed although more likely the cells will be descendants of the transformed cells) under conditions for expression of the gene, so that the encoded polypeptide is produced. If the polypeptide is expressed coupled to an appropriate signal leader peptide it may be secreted from the cell into the culture medium. Following production by expression, a polypeptide may be isolated and/or purified from the host cell and/or culture medium, as the case may be, and subsequently used as desired, e.g. in the formulation of a composition which may include one or more additional components, such as a

pharmaceutical composition which includes one or more pharmaceutically acceptable excipients, vehicles or carriers.

Further aspects and embodiments of the present invention 5 will be apparent to those skilled in the art, in view of the present disclosure. To facilitate understanding of various aspects of the invention the following explanation of the inventors' work and how it is applicable in the present invention is provided, supplementing the experimental work 10 described in further detail later.

Classification of α/β -Barrels

The basic α/β -barrel framework consists of at least 200 residues arranged in eight parallel β -strands connected and 15 surrounded by eight helices, with a central hydrophobic core. Anyone familiar with protein structure can identify the strands and helices by inspection of molecular models or by use of computer programs such as Rasmol (<http://www.mrc.cpe.cam.ac.uk/cpe/manuals/ccp4/rasmol.html>), 20 Molscript (Kraulis et al. *Biochemistry*, 1994, 33: 3515-3531), CATH or SCOP. The barrel structure can sometimes be circularly permuted by connecting the N and C-termini and cutting elsewhere by changing the DNA that codes for the protein. However, someone skilled in the art will know 25 where the original N-terminus would have been. The numbering herein of the sequence of strands and helices is based on the conventional position of the N-terminus. The strands in the barrel are numbered sequentially β_1 to β_8 and the helices α_1 to α_8 from the N-terminus. These are arranged 30 such that strand β_8 is adjacent to and hydrogen-bonded with strand β_1 . In a few cases, the barrels do not have eight

parallel β strands. There are barrels that contain ten parallel β strands.

5 The active site is always in the same region of the protein, at the C-terminus, and is formed by residues of the eight loops connecting the carboxy end of each strand with the amino end of the following helix.

10 The α/β -barrel enzymes have two sets of loops. The C-terminal end contains a β -loop- α unit, which presents wide variation in their structure and length. The loops in the α -loop- β units within the barrel, are shorter and they can adopt two different conformations for strand entry into the parallel β sheet. Bränden, C., *supra*. Chothia, C. & Lesk, A. M. Conformations for strand entry into parallel β sheets pp49-58 (1991). *In Molecular Conformation and Biological Interactions*. Ed. Balaram P and Ramaseshan, S. Indian Academy of Sciences, Bangalore.

20 In the scaffold there are mainly three pieces that can be combined and exchanged: the lid of the active site (variable region); the hydrophobic area and the charged area in the binding site (constant region) see below.

25 As noted, the active site lid is the structure that covers the active site, closing and shielding it from solvent. It may consist of or comprise loops at the carboxyl termini of the β -strands (e.g. $\beta 1\alpha 1$, $\beta 6\alpha 6$), extra N-terminal segment, extra domains (between $\beta 3\alpha 3$) and/or C-terminal segment (after $\beta 8$). More of 70 % of catalytic residues in the α/β -barrel enzymes appear in these structural motif. These residues are directed involved in the rate-limiting

step in the reaction mechanism. The rest of the catalytic residues are located in the loops at the carboxyl termini that form the binding site. They are involved in specific substrate binding and catalysis, but their main role is 5 interaction with the substrate (holding it in the correct position), and they do not participate in the rate-limiting step in the reaction mechanism.

The binding site is the structure (mainly loops) at the 10 carboxyl termini of the β -strands that form the funnel-shaped pocket and contain 90% of the residues that participate in binding (holding the substrate in the correct position for the catalysis) and 30 % of residues that participate in binding and catalysis in the overall reaction 15 but not in the rate-limiting step reaction mechanism.

The binding site can be divided in two areas, on the basis of the chemical nature of amino acid side-chains which form it. There is a hydrophobic area and a charged area. The residues in the hydrophobic area are more than 60% 20 hydrophobic residues (e.g. leucine, isoleucine, alanine, valine, phenylalanine). The residues in the charged area are more than 60% positive, negative or polar amino acid residues (e.g. aspartic, glutamic (-), lysine, arginine (+), asparagine, glutamine, cysteine, histidine, tryptophan).

25 Fersht, *supra* . Bränden, C., *supra*.

Since the localisation of the hydrophobic and the charged 30 binding sites remain constant, they may be considered as "constant" pieces. Mutation of the constant pieces may be used to change substrate binding. Among the constant features, there is a variable region, the "covering lid" placed over the site, which closes and shields it from the solvent.

The "constant" pieces:

Phosphate or charged binding site.

5 The constant region: e.g. β 7- α 7, β 8- α 8 segments are part of the phosphate-binding site in at least 10 different α / β -barrels. Farber & Petsko *TIBS* 15, 228-234 (1990). Reardon & Farber *FASEB J.* 9, 497-503 (1995). Wilmanns et al. *Biochemistry* 30, 9161-9169 (1991); Bränden, C., *supra*.

10 Small modifications in these "constant" regions cause different orientations of the phosphate group of the substrate which may lead to changes in substrate affinity, e.g. those with PRAI and IGPS Wilmanns, M., Priestle, J. P., Niermann, T. & Jansonius, J. N. Three-dimensional structure

15 of the bifunctional enzyme phosphoribosylanthranilate isomerase: indoleglycerolphosphate synthase from *Escherichia coli* refined at 2.0 Å resolution. *Journal Of Molecular Biology* 223, 477-507 (1992).

20 Hydrophobic pocket.

The β 2- α 2 and β 4- α 4 are part of the hydrophobic pocket in the active site. For glycolate oxidase and flavocytochrome b_2 , a few mutations in the active site have been fine-tuned to make them effective on different substrates. Bränden, C.,

25 *supra*.

"Variable region" Active site lid.

30 *Extra N-terminal segment.* The N-terminal structural segment that is not part of the α / β barrel and leads into strand β 1 Bränden, C., and Tooze, J. *supra*

β_1 - α_1 loop. The structure at the carboxyl termini of the β -strand number 1 that leads in the α -helix 1. Bränden, C., and Tooze, J. *supra*.

5 β_6 - α_6 loop. The structure at the carboxyl termini of the β -strand number 6 that lead in the α -helix 6. Bränden, C., and Tooze, J. *supra*

10 *Metal binding site.* In some superfamilies (e.g. metal-dependent hydrolases) the structural segments β_5 - α_5 and β_7 - α_7 , together with the C-terminus, are part of the metal-binding site. Bränden, C., and Tooze, J. *supra*

15 *Loops forming others domains.* An additional loop region from a second domain or a different subunit may comes close to the active site and participate in binding and catalysis, as is found for pyruvate kinase and amylase in which the loop β_3 - α_3 is folded in a separate domain. Bränden, C., and Tooze, J. *supra*

20 *C-Terminus segment.* The segment from the C-end of barrel which presents wide variation in its structure and length (Table I and II). It is considered as starting from the C-end of β_8 . In some enzymes, the C-end is part of the lid.

25 Bränden, C., and Tooze, J. *supra*

30 The classification devised by the present inventors is based on the structures of phosphoribosylanthranilate isomerase (PRAI) and indole-3-glycerol-phosphate synthase (IGPS) as models (Table I and Table II).

The main structural feature of the active site lid in the Class I (or PRAI group) of α/β -barrel proteins is mainly the connection β 6- α 6 (10-12 residues), which is rich in glycine residues. For example, PRAI, triosephosphate isomerase, 5 class II aldolases and pyruvate kinase, which belong to this first class, contains the highly conserved sequences GXGGXG, GXG or GXXG. The lack of side chains in the loop β 6- α 6 is sterically favourable to its approaching to the remainder of the structure and thus covering the active site. We call 10 this Class I or "lotus leaf" lid (Table I and Figure 1). The class I group is characterised by the absence of an N-terminal extension, or its replacement by a very short segment (2-9 amino-acid residues), generally accompanied by a characteristically short β 1- α 1 connection segment (2-11 15 residues).

The IGPS domain belongs to Class II (Table II and Figure 1). Its lid is shaped as a clover leaf and encompasses three main substructures. The first two structural segments 20 present wide variations in their structure and length. These are an extra N-terminal segment, and β 1- α 1 structural segment. The number of residues in both components together varies from 18-89 residues (Table I and II). The segment connecting β 6 to α 6 (10-12 amino acid residues) does not 25 contain any particularly conserved sequence among different superfamilies. It is positioned to interact with the N-terminal segment when the lid is closed over the binding site.

30 *Correlation between the structural class of the lid and the reaction mechanism.*

The structure of the active site lid relates to the mechanism (Table III). For example, triosephosphate isomerase and xylose isomerase both catalyse aldose-ketose isomerisations of different substrates. The first enzyme 5 belongs to class I and uses a proton-transfer mechanism.

The second one (Class II) has a hydride transfer mechanism.

In an enzyme family that catalyses the same reaction by the same mechanism but for different substrates, the classification of the lid remains the same, but the lids 10 vary in length and sequence to generate the different specificities (Table III). For example, aldol-ketol isomerisations in TIM-like aldol-ketol isomerasers are mechanistically related to 2-hydroxyaldimine-ketoamine isomerisations (a reaction known as Amadori rearrangement) 15 in PRAI. In both cases, general-base catalysed proton abstraction and repositioning occur, although the reaction intermediates are different. Both enzymes belong to class I (Table I and III). The metal-dependent hydrolase superfamily is another example of this. This family uses a dozen 20 different substrates and is responsible for seven of some 20 steps along four important metabolic pathways. They have a common reaction mechanism; the metal ion (or ions) activates a water molecule for nucleophilic attack to the substrate. They are all in our Class II (Tables II and III). 25

Variations on the C-terminal regions of the barrel and the active loop regions. Changes in residue spacing plays a major role in evolution of protein function, with insertions and deletions contributing substantially to the 30 diversification of enzyme activities. At one level in the α/β -barrel family, such changes can lead to changes in specificity although retaining membership of class I or II. An interesting example is the enolase superfamily (Class

II). During evolution they have retained the structural strategy of catalysing the chemically difficult step of α -proton abstraction but they gained additional functional groups to catalyse different overall reactions. Further, 5 more radical changes can lead to the change of lid design, accompanied by a change in class and a change in mechanism or evolve new function e.g. those with PRAI and IGPS.

See Annex 1 below for further discussion of evolution of new 10 catalytic activities.

As described in the experimental section below, the inventors proved the principle of the invention by converting an α/β -barrel protein indoleglycerolphosphate 15 synthase (IGPS) into phosphoribosylanthranilate isomerase (PRAI). The resultant enzyme is of similar catalytic activity to the naturally occurring enzyme, and, at low substrate concentrations, is even more active.

20 *Combinatorial design*

The invention thus provides a general procedure for producing new enzymes, employing what may be termed combinatorial design.

25 The invention generally provides for design and production of an enzyme that catalyses a desired reaction on a desired, or target, substrate.

In one approach according to the invention, a barrel binding the desired substrate is selected or provided, either by 30 choosing a naturally occurring barrel which binds the substrate or by mutating and selecting another barrel. Such selection will generally involve determining ability of a barrel to bind the target substrate, and may employ any

technique available in the art, for instance phage or ribosome display. See e.g. Fersht, *supra*, chapter 14.

A lid, based on the template of a lid for an α/β -barrel that 5 catalyses the desired reaction or a reaction of the desired type, is grafted on to or engineered into the barrel that binds the substrate, to combine a binding site for the target substrate with a catalytic template.

10 The lid is then subjected to targeted mutation and selection. Rules and guidance for this are provided below.

Both lid and substrate binding sites may be subjected to mutation and selection to alter or optimise respective 15 properties, e.g. one or more of binding affinity and catalytic activity.

Transplantation of Class I and Class II lids

Examination of the classes identified herein leads to 20 recognition of where the catalytic groups are and so which should be or should preferably be transplanted.

Summary of location of catalytic loops

25 Class I

Catalytic groups are mainly in the $\beta_6-\alpha_6$ loops; some catalytic groups are in $\beta_1-\alpha_1$ and $\beta_2-\alpha_2$ loops. (The $\beta_6-\alpha_6$ loop connects strand β_6 and helix α_1 , etc).

30 Class II

Catalytic groups are mainly in the $\beta_1-\alpha_1$ and $\beta_6-\alpha_6$ connecting loops and the N-terminal extension and C-terminal extension.

A Class I lid that catalyses a particular reaction may be grafted onto a Class II scaffold as follows:

- the N-terminal extension of the Class II scaffold is deleted;
- the β 1- α 1 loop is shortened;
- the β 6- α 6 loop is modified.

A Class II lid that catalyses a particular reaction may be

10 grafted onto a Class I scaffold as follows:

- an N-terminal extension is added;
- the β 1- α 1 loop is lengthened;
- the β 6- α 6 loop is modified.

15 *Choice of substitutions*

Loops may be changed to a consensus sequence found from examining a family of α / β -barrels that catalyse the desired reactions.

20 More detailed practical points to consider

1. **Choice of scaffold for the desired function or catalytic activity**

25 The suitable scaffold is chosen, and this may take into account biochemical and structural analysis, considering any one or more of the following:

Biochemical data for scaffold and reference proteins

30

- a) Is the scaffold a monomeric or an oligomeric protein? A monomeric protein may be preferred, where available.

- b) Is there a good expression level in bacteria and is it a well-characterised gene? Is it part of a regulon? Is it part of a metabolic pathway? Can we use *in vivo* selection?.
- c) What is known about its function, activity assay, ligands (substrates, inhibitors, effectors, metals, etc)
- 5 d) Kinetic characterisation: kinetic parameters, kinetic mechanism.
- e) Reaction mechanism.
- f) Role of specific residues from mutagenesis studies
- 10 g) Molecular properties in solution.
- h) Folding studies.

Structural data of both scaffold and reference proteins.

- 15 a) Primary structure. Sequence alignment, identification of orthologous proteins (proteins with the same activity in different species), neighbour families (proteins with conserved structural or functional patterns). Consensus sequences. Conserved signatures.

20

Secondary structure: α/β -barrel family fold.

- b) 3-D structure of enzyme-ligand complex, apoprotein and/or holoprotein structure. Detailed description of the active site: lid, the binding site and the topology of the molecule.
- 25 3-D analysis of both proteins, using the PDB (Protein Data bank),
CATH (see above) (Thorton, *supra*), FSSP (Fold classification
- 30 based on Structure-Structure alignment of Proteins) (L. Holm and C. Sander. Mapping the protein universe. *Science* 273:595-602 (1996). The FSSP database is based on exhaustive all-against-all 3D structure comparison of protein

structures currently in the Protein Data Bank (PDB). The classification and alignments are automatically maintained and continuously updated using the Dali search engine. See more details in <http://www2.ebi.ac.uk/dali/fssp/>.

5

2. Design

The following provides guidance for embodiments of the present invention.

10

a) Active site lid. Based on the active site lid classification provided herein, firstly identify the class to which the lid of the desired protein belongs. How many components are part of the lid? a practical rule consists 15 in focusing on the N-extra terminal segment and the loops $\beta_1-\alpha_1$, $\beta_6-\alpha_6$, $\beta_3\alpha_3$ (looking for extra-domains). When fragments of the loops $\beta_7-\alpha_7$, $\beta_5-\alpha_5$ are part of the lid, this means that the template of the metal binding site is involved in catalysis. Use CATH database to get the 20 topology of your protein. See more details in <http://tops.ebi.ac.uk/tops/ExplainDetailed.html>

In general, there is a correlation between the length of the N-extra amino terminal segment and the length of the loop 25 $\beta_1-\alpha_1$, i.e., both are short or long. In the lid class I, the leading structural feature of the active site lid is the connection $\beta_6-\alpha_6$, which is rich in glycine residues. In the lid class II there are at least three main components: N-extra amino terminal segment, the loop $\alpha_1-\beta_1$, the loop $\beta_6-\alpha_6$ and sometimes in addition there are fragments of loops 30 $\beta_3-\alpha_3$, $\beta_7-\alpha_7$, $\beta_5-\alpha_5$ or the C-end segment. The next step may be identification of the residues involved in catalysis,

which are usually localised in the lid (Altamirano and Fersht, *supra*). Further, the lid plays an additional role in substrate discrimination because the size of the ligands is related to the class lid (Fersht *supra*). Finally, the 5 conserved features in the lid within different members of the family (Altamirano, et. al, submitted) may be identified using FSSP program (see above)

b) Constant regions. Attention focuses on the binding site. 10 We identify the polar region and the most hydrophobic area in the active site. The polar region commonly appears localised between the loops β 7- α 7 and β 8- α 8 (the phosphate binding site) while the hydrophobic area is localised between β 2- α 2, β 4- α 4, β 3- α 3.

15 We identify the residues directly involved in substrate interaction and the conserved features within different members of the family. Now we are ready to the superposition of both structures and other neighbouring structures. See 20 the next section.

Thus, the binding site of the α / β -barrel family can be divided into three regions in order to locate and modify the sections of the protein involved in catalysis and binding in 25 accordance with the present invention:

1. *Active site lid - the primary determinant of the chemical reaction that is catalysed*

This consists of the loops β 1 α 1 and β 6 α 6, the extra N-terminal region and the carboxyl terminus. The lids are 30 divided into two classes.

Class 1 lid: β 1 α 1, the extra N-terminal region and the carboxyl terminus are characteristically shorter than Class

2 lid components. The β 6 α 6 loop often has a distinctive sequence composition that is rich in glycine residues.

Class 2 lid: β 1 α 1, the extra N-terminal region and the carboxyl terminus are characteristically longer than Class 1 lids. The β 6 α 6 loop tends to be longer than a class 1 lid component. Class 2 lids are more abundant, and their structures are more adaptable.

The active site lid dictates the nature of the reaction catalysed.

10 2. *Hinge region*

The hinge region consists of the last two residues of each β -strand and the first two residues of each of the associated loops. They can have residues that are involved in catalysis and binding.

15 3. *Body loops - important in specificity*

Loops β 2 α 2 and β 4 α 4 bind the hydrophobic regions of the substrate.

Loops β 7 α 7 and β 8 α 8 bind the charged regions of the substrate.

20 Strands β 3, β 5 and β 8 can contain the metal binding sites.

Loop β 3 α 3 may be recruited into the hydrophobic binding site.

3. *3-D superposition analysis.*

25

First, focus attention on the scaffold. Note the barrel shape and the segments having a counterpart in the other protein. The next step consists in the identification of segments that may overlap, with a r.m.s.d. of 2-3 Å.

30

Finally, focus attention on the segments with more than five residues that cannot be structurally aligned. Identify all

the insertions and deletions or any other drastic changes in the secondary structure. Identify the segments that can be aligned by joining each insertion or deletion. Use FSSP database (see supra)

5

1 Select the segments that have more than five residues that cannot be structurally aligned and use them as target points.

10 Analyse these structural data on the light of the study of their functions. Is the active-site lid the target? If so, then use data about reaction mechanism, catalytic residues, the structural components of the lid and the lid class. Is the binding site part of the "constant regions" in

15 the target? If so, use data about interaction with ligands, affinity constants, stereochemical constraints, etc.

4. Design the modification of scaffold.

20 Select the segments on which you will graft the lid or the binding site by insertion, deletion or random target mutagenesis. Concentrate on the segments chosen as pivots (joint points) of the segment or the segment to be deleted.

25 To make an insertion, choose the random mutation carefully and the conserved sequence, introduce the superfamily consensus sequences (Conserved residues among the evolution in different species). Design a set of synthetic DNA fragments of the target points from diverse species. The

30 scaffold is now ready for fitting its shape and its function.

Outline of a procedure in accordance with the present invention

Step 1

5 *Provision of a scaffold including binding site for substrate*

Case A

If a known α/β -barrel has the desired binding site for the substrate, then employ this. In this case, a lid will be 10 chosen from another α/β -barrel that catalyses the desired reaction, or a similar reaction, one of the same type.

Case B

If there is no known α/β -barrel with the desired binding 15 site for the substrate, then a scaffold is chosen that catalyses the desired reaction with a similar substrate. That is, a scaffold is chosen that catalyses the desired reaction and has some features in its binding site that may be adjusted for binding the desired substrate (e.g. its 20 hydrophobic or charged regions). In this case the scaffold will be mutated (see below) and a variant which binds the substrate will be selected (see below).

Step 2

25 *Selection of targets for mutagenesis from superposition of 3-D structures.*

There are two major components in the scaffold; one mainly for the binding site and one mainly for the reaction 30 mechanism. There are three regions that can be modified: the hydrophobic and the polar parts of the substrate binding site; and the catalytic lid.

Case A

(Where a scaffold for binding the substrate is known in a protein, and there is another protein known that catalyses the desired or a similar reaction).

5

The substrate binding scaffold is used and an appropriate template catalytic lid is grafted on.

10 For the choice of lid for the reaction mechanism, conserved features in the superfamily may be examined and superposed with those of the binding scaffold.

Case B

15 (Where a protein is known that catalyses the appropriate reaction on a different substrate that has similarities to the desired substrate.)

20 For the binding site, conserved features in the superfamily may be examined, and superposed with those of the binding scaffold.

25 In either of Case A and Case B, target residues for mutagenesis (by insertion, deletion or introduction of consensus sequences) may be chosen as segments of five or more residues that can not be structurally aligned with the consensus of those from the superfamily.

Step 3*Mutagenesis and selection*

30 Convenient methods for mutagenesis, sexual recombination and selection of active protein are available in the art, and some are described below. These generally involve design and preparation of synthetic DNA fragments for creating

further diversity in the target sequences. The shape of the barrel may be refined for improving its function by *in vitro* evolution methods.

5 Each of Cases A and B have been exemplified experimentally, as describe in more detail below. Further, brief details and discussion are provided here.

CASE A

10 (Use of a preexisting binding site and grafting a template active-site lid, which is modified by insertions, deletions and/or recombination).

Step 1

15 *Scaffold selection*

A monomeric α/β -barrel protein, the indole-3-glycerol-phosphate synthase (IGPS), was chosen as a scaffold able to bind the desired substrate.

20 The desired enzyme activity was that of phosphoribosyl anthranilate isomerase (PRAI).

Selection system

An *in vivo* selection strategy for PRAI activity was designed 25 based on complementation of *E. coli* JA300 (a PRAI-deficient strain that does not grow in the absence of tryptophan (Trp)). In *E. coli*, PRAI and IGPS are part of the same 45 kDa polypeptide chain specified by the *trpC* gene. However, *E. coli* JA300 carries the W3110 (*trpC1117*) allele and so 30 lacks isomerase activity, but retains normal levels of synthase activity. Complementation provides indication that the specific clone contains a plasmid expressing an IGPS variant with PRAI activity.

Step 2*3D superposition*

5 The structures of IGPS and PRAI were superimposed using the program SETOR

Scaffold

10 All β -strand residues of the central β -barrel of PRAI have counterparts in the IGPS. The 68% of the α -helical residues have structurally equivalent residues in the other domain.

Active site lid class

15 The IGPS active site is covered by the N-terminal α 0 helix, and by the β 1- α 1 (15 residues), β 2- α 2 (9 residues) and β 6- α 6 (11 residues) loops, all located at the C-terminal side of the α / β -barrel. This defines the IGPS protein as having a class II active site lid.

20 PRAI, however, has an very different active site lid that is mainly formed by the β 2- α 2 (10 residues), β 6- α 6 (11 residues) and β 8- α 8 (12 residues) loops. PRAI has a class I active site lid.

25 *Constant regions in the active site*

The β 2- α 2 loop in both enzymes is involved in binding the anthranilate moiety of the respective substrates PRA and CdRP. The β 8- α 8 loop comprises the phosphate binding site. The superposition of the two structures reveals almost identical locations but different orientations of the phosphate binding site. Since both loops (β 2- α 2, β 7- α 7 and β 8- α 8) are similarly arranged in the two enzymes, the target

of our selection was solely the extra N-terminal end (helix α_0 and two bends), the β_1 - α_1 loops and the β_6 - α_6 loops.

Active site lid as the target for switching reaction

5 *mechanism*

The first step was grafting a PRAI lid on to a IGPS scaffold that contains a common binding site. The process included the deletion of 48 amino acid residues from the amino terminal end of IGPS; this deletion mutant was called 10 (IGPS49). The IGPS49 scaffold was further modified by replacing 15 amino acid residues corresponding to the β_1 - α_1 loop by a new randomised segments of 4 to 7 amino acid residues. The gene encoding IGPS49 was used as template to create three new libraries IGPS49L1 (GKXXG), IGPS49L1RGD 15 (GKXRGD) and IGPS49L1SV (length size variation: GKXX, GKXXX, GKXXXX or GKXXXXX) via PCR methodologies including overlap extension PCR, inverse PCR and random primer PCR.

20 The next set of modifications involved the $\beta_6\alpha_6$ loop, including the introduction of an aspartic residue at position 184 (acting as a general base in the active site) and also the PRAI consensus sequence GXGGXGQ21, with the aim 25 of improving the active site lid. A new library called IGPS49L1L6 was constructed using the IGPS49L1, IGPS49L1RGD and IGPS49L1SV libraries as templates.

In vitro recombination to improve the fit of the barrel shape and its function followed by in vivo selection

30 In this phase, a first round of DNA shuffling was performed with the pool of genes from the selected clones that were able to grow at very low concentration of Trp. A second round of recombination was performed by DNA shuffling and Staggered extension procedure (StEP), using the pool of 80

colonies selected from the first round and synthetic DNA fragments encoding for the protein segments corresponding to loops β 1 α 1, β 6 α 6, β 4 α 4 from diverse species of PRAI. The *in vivo* selection yielded 360 colonies capable of growing in
5 the absence of any exogenous Trp.

In vitro-evolved PRAI

The newly evolved phosphoribosylanthranilate isomerase has similar catalytic properties to the natural enzyme, with an
10 even higher specificity constant.

CASE B

A scaffold containing a catalytic lid was selected and changes made in the binding site (constant pieces).

15

Step 1

Scaffold selection

Human Phosphotriesterase homology protein (PHP) was chosen as a scaffold. It binds the substrate for the desired
20 enzymatic activities.

The desired enzymatic activities were phosphotriesterase (PTE) activity and phosphodiesterase (PDE) activity.

25 PHP does not have a known enzymatic activity, though it has 28% sequence identity with phosphotriesterase, is monomeric and binds two zinc ions per monomer. Unlike phosphotriesterase, PHP does not catalyse either the hydrolysis of nonspecific phosphotriesters or
30 phosphodiesters (promiscuous activity in PTE). Phosphotriesterase is an enzyme capable of hydrolysing both widely employed pesticides and phosphofluoridates.

*Step 2**Sequence Alignment*

5 Phosphotriesterase, PHP (*E. coli*), PHP (*M. pneumoniae*), PHP
5 (*M. Tuberculosis*), PHP (mouse) and PHP (human) are 27-30%
10 identical in amino acid sequence. The aspartate and all
10 four histidine residues that coordinate Zn²⁺ in
10 phosphotriesterase are conserved across the six PHP
10 proteins. Only the carbamylated lysine at position 169 is
10 not strictly conserved. This residue is replaced by a
10 glutamate and is shifted by one position in the alignment
10 for ePHP, muPHP, rPHP, hPHP .

3-D superposition

15 The structures of PHP from *E. coli* and PTE were superimposed
15 using the program DALI.

Scaffold

20 All β-strand residues of the central β-barrel of PHP have
20 counterparts in the PTE. More than 70% of the α-helical
20 residues have structurally equivalent residues in the other
20 domain.

Active site lid class

25 The PTE active site is covered by the N-terminal (residues
25 35-51, including two strands of antiparallel β-sheet), the
25 β1-α1 (residues 56-76, including β-sheet, turns and helical
25 turn), β6-α6 (residues 229-237) loops and a segment of β7-α7
25 (only residues 254-256) all located at the C-terminal side
30 of the β-barrel. The lid class is II.

PHP has slightly different active site lid, the N-terminal segment is shorter (8 residues). The lid is mainly formed by the β 1- α 1 (18 residues, that encompasses antiparallel β -strands, residues 17-32), β 6- α 6 (11 residues, is quite 5 similar in both proteins). PHP has class II active site lid.

Constant regions in the active site

Importantly, significant differences between the two structures are found in the regions corresponding to the 10 binding site of the PHP. The β 3- α 3 loop is involved in binding the substrate with hydrophobic and smaller leaving groups such as ethoxy groups in both proteins. In PTE, the β 7- α 7 loop has an insertion of 14 residues, and the β 8- α 8 loop has an insertion of 8 residues with respect to the PHP 15 sequence. These bind the phosphorus centre and are involved in binding the substrate with hydrophobic large and bulkier leaving group such methylbenzyl group. The superposition of the two structures reveals almost identical locations for the residues involved in metal ligation. Since the lids 20 including the metal binding site are similarly arranged in the two enzymes, the target of the selection were a fragment of the loop β 7- α 7 (residues 260-276) and all the β 8- α 8 loop.

Constant pieces as the target for switching specificity

25 The first step in the design was grafting a template of the PTE substrate binding site on to a PHP scaffold by insertion of 18 amino acid residues in the loop β 7- α 7 of PHP. The PHP (+ 18 residues) scaffold is further modified by inserting 8 amino acid residues corresponding to the β 8- α 8 loop by a new 30 randomised segments via PCR methodologies including overlap extension PCR, inverse PCR and random primer PCR.

The binding depends significantly on the relative size and orientation of the two subsites that accommodate the coordination of the alkyl or aryl substituents within the enzyme active site. Using *in vitro* evolution methods, the 5 present invention enables redesign of the active site to alter and enhance the substrate specificity of the new evolved PTE.

Evolving phosphodiesterase activity in PHP.

10 The full negative charge within the phosphodiester substrate is thought to be primarily responsible for the slow rate of catalytic hydrolysis of these compounds by the PTE. The active site of the PTE is largely hydrophobic, and thus it would not be expected to accommodate the negative charge on 15 the substrate very well. Further, the nucleophile in the active site (metal-bound hydroxide ion) may not be able to attack the anionic substrate effectively. In order to evolve phosphodiesterase activity, we include a set of modifications: the insertion of the IGPS phosphate binding 20 site corresponding to the β 7 α 7 and β 8- α 8 loop. This new binding site is able to accommodate the negative charge on the substrate.

25 *In vitro recombination to improve the fit of the barrel shape and its function followed by selection, in vivo or in vitro.*

30 The *in vivo* screening system employs expression of the protein in the periplasm and using the strong yellow colour or display strong fluorescence produced by the hydrolysis of the substrate (Paroxon or Diisopropyl fluorophosphate). The clones with PTE activity become yellow or with fluorescence.

*Summary of primary grafting rules*Active site lid - to direct the chemical mechanism

1. The sizes and composition of the lid components ($\beta 1\alpha 1$ and $\beta 6\alpha 6$, the extra N-terminal region and the carboxyl terminus) are grouped according to the Class 1 or Class 2 size and composition categories.
2. The sequences of the lid components in orthologous enzymes that catalyse the desired reaction are examined and consensus sequences or conserved residues identified to be included in the template loops that are transplanted.
3. The size of the cavity covered by the lid may be increased or decreased by altering the sizes of the side chains.

Hinges

- 15 The hinge regions of loops may be included with the loops that are transplanted into the scaffold because they may have important residues.

Body loops - to tailor the substrate specificity

1. The sequences of body loops of known α/β -barrel proteins that bind the desired substrate are examined and consensus sequences and residues included in the loops to be transplanted.
2. If the desired substrate is not bound by other known enzymes, then the proteins that bind the closest examples are preferably used as models. The modifications to the loops to accommodate the substrate can be based on the size of the hydrophobic and charged moieties of the desired substrate relative to known examples using the principle that loops $\beta 2\alpha 2$ and $\beta 4\alpha 4$ bind the hydrophobic regions of the substrate and $\beta 7\alpha 7$ and $\beta 8\alpha 8$ bind the charged region. The body loops may also be tailored to accommodate polar substrate residues in the hydrophobic site and hydrophobic

residues in the charged site. The size of the hydrophobic site may increased or decreased according to the size of the substrate. Modifications may be made to loop $\alpha 3\beta 3$ to compensate for changes in the size of $\beta 2\alpha 2$ and $\beta 4\alpha 4$.

5 3. If the substrate is greatly different from any known example, then substructures of the substrate may be identified (e.g., aromatic rings, nucleosides, sugar rings, phosphate groups or aliphatic side chains) and then the loops from known proteins that bind these substructures can
10 be recruited. It is most useful to choose proteins that bind more than one of these substructures simultaneously.

Creation of diversity

It is desirable to create diversity in the loops and segments that are grafted by using deletions and insertions
15 and substitutions of sequences that can be found from examination of naturally occurring in orthologous families.

EXPERIMENTAL

20 *SECTION 1*

*CLASSIFICATION OF α/β -BARREL PROTEIN LIDS AND IMPLICATIONS
FOR ENZYME DESIGN*

25 In this first section, combinatorial design principles in α/β -barrel proteins for the creation of novel biocatalysts are described.

30 The α/β -barrel motif is Nature's favourite fold for the generation of enzymatic activity. Nature appears to have evolved a structural framework enabling the rapid evolution of active sites, the understanding of which facilitates the design of new proteins *in vitro*. There are two constant

features in the active sites of many α/β -barrels, which differ only in detail: a binding site for the phosphate or any other charged group of the substrate; and a hydrophobic binding site. Mutation of these lead to changes in substrate binding. Between the constant features are variable regions that contain most of the catalytic residues, the "covering lids". The inventors here categorise the α/β -barrel domains into two classes, according to the overall template structure of the lids, and indicate that the template of the lid dictates the type of reaction mechanism. The combinatorial association of lids and constant binding regions coupled with mutation and selection provides a basis for generation of new enzymatic activities *in vitro*, as is proven in the experimental example in Section 2 below.

15

The α/β or TIM (triosephosphate isomerase) barrel is the most common motif in enzyme structure and is the basic scaffold of enzymes catalysing a wide variety of reactions (Farber & Petsko *TIBS* 15, 228-234 (1990); Murzin et al. *J. Mol. Biol.* 247, 536-540 (1995); Reardon & Farber *FASEB J.* 9, 497-503 (1995); Holm & Sander *Nucleic Acids Res.* 24, 206-209 (1996); Chothia & Lesk Conformations for strand entry into parallel β -sheets pp49-58 (1991). In *Molecular Conformation and Biological Interactions*. Ed. Balaram P and 25 Ramaseshan, S. Indian Academy of Sciences, Bangalore'. The basic framework consists of at least 200 residues arranged in eight parallel β -strands connected and surrounded by eight helices, with a central hydrophobic core. The α/β -barrel enzymes have a variety of quaternary arrangements and 30 show little or no homology, except for those that catalyse the same reactions in different organisms. Nevertheless, their active site is always in the same region of the

protein, at the C-terminus, and is formed by the eight loops connecting the carboxy end of each strand with the amino end of the following helix (Lesk et al. *Proteins* 5, 139-148 (1989); Murzin et al. *J. Mol. Biol.* 236, 1382-1400 (1994); Murzin et al. *J. Mol. Biol.* 236, 1369-1381 (1994)¹). The connections between the C-termini of helices and strands usually involves short loops, whereas those from strands to N-termini are long and provide a structural basis for binding and catalytic sites. Most of the catalytic residues in the α/β -barrel enzymes appear in these loops, which form a covering lid over the site, shielding it from solvent (Bränden *Curr. Opin. Struct. Biol.* 1, 978-983 (1991); Bränden and Tooze *Introduction to Protein Structure*, 2nd. Edition (Garland Publishing Inc., New York, 1999)¹. There are two constant features in the barrel: a hydrophobic region that binds part of the substrate; and a phosphate binding site, which may be modified to bind other charged groups, such as metal ions. The α/β -barrel fold has been extensively analysed from an evolutionary perspective. Farber et al., (Farber & Petsko *TIBS* 15, 228-234 (1990); Reardon & Farber *FASEB J.* 9, 497-503 (1995)¹, based on mainly on structural criteria but also on function divided the α/β -barrel proteins into six structural families (A-F). Chothia and colleagues noted that the pattern of packing inside the β -barrel of glycolate oxidase and ribulose-1,5-biphosphate carboxylase oxygenase (rubisco) is similar and differs from that inside the barrel of triosephosphate isomerase, which has the most asymmetric cross section and is very distorted (Lesk et al. *Proteins* 5, 139-148 (1989); Murzin et al. *J. Mol. Biol.* 236, 1382-1400 (1994); Murzin et al. *J. Mol. Biol.* 236, 1369-1381 (1994)¹. Petsko (Neidhart et al. *Nature* 347, 692-694 (1990); Neidhart et al. *Biochemical Society Symposia* 57, 135-141 (1990)¹ and Bränden

(Bränden *Curr. Opin. Struct. Biol.* 1, 978-983 (1991); Bränden and Tooze *Introduction to Protein Structure*, 2nd. Edition (Garland Publishing Inc., New York, 1999)' analysed two sets of evolutionary related enzymes that perform 5 different biological functions (mandelate racemase and muconate lactonising enzyme, and glycolate oxidase, flavocytochrome b_2 and mandelate dehydrogenase). They suggested that the proteins could have evolved by divergent evolution by retaining the chemical mechanism but with 10 mutations in the barrel leading to different specificities.

From a survey of the α/β -barrel domains in the SCOP, CATCH and Dali databases (Murzin et al. *J. Mol. Biol.* 247, 536-540 (1995); Orengo et al. *Structure* 5, 1093-1108 (1997); 15 Holm & Sander *Nucleic Acids Res.* 22, 3600-3609 (1994); Holm & Sander *TIBS* 20, 478-480 (1995); Hubbard et al. *Acta Crystallogr D Biol Crystallogr* 54, 1147-1154 (1998)) the inventors now provide two broad classes into which these 20 proteins can be categorised according to the structural design of the covering lid. The lid contains most of the catalytic residues, and so understanding the design of the lid is a key step in designing novel activities based on the α/β -barrel scaffold. In particular, this allows for 25 mutation, recombination and alteration of the lid while retaining a substrate binding site, thereby altering the reaction catalysed by the enzyme on the bound substrate.

The classification is based on the structures of phosphoribosylanthranilate isomerase (PRAI) and indole-3-glycerol-phosphate synthase (IGPS) as models (Table I and 30 Table II, Figure 1), as has been described already above.

The inventors have realised that the structure of the active site lid appears to dictate the type of reaction mechanism (Table III). For example, triosephosphate isomerase and xylose isomerase both catalyse aldose-ketose isomerisations of different substrates (Banerjee et al. *Protein Engineering* 8, 1189-1195 (1995); Farber et al. *Biochemistry* 28, 7289-7297 (1989)'). The first enzyme belongs to class I and uses a proton-transfer mechanism. The second one (Class II) has a hydride transfer mechanism. In an enzyme family that 5 catalyses the same reaction by the same mechanism but for different substrates, the classification of the lid remains the same, but the lids vary in length and sequence to generate the different specificities (Table III). For 10 example, aldol-ketol isomerisations in TIM-like aldol-ketol isomerases are mechanistically similar to 2-hydroxyaldimine-ketoamine isomerisations (the Amadori 15 rearrangement) in PRAI. In both cases, general-base catalysed proton abstraction and repositioning occur, although the reaction intermediates are different. Both 20 enzymes belong to class I (Table I and III). The metal-dependent hydrolase superfamily is another example of this (Gerlt & Babbitt et al. *Curr Opin Chem Biol* 2, 607-612 (1998)). This family uses a dozen different substrates and is responsible for seven of some 20 steps along four important 25 metabolic pathways (Holm & Sander *Proteins* 28, 72-82 (1997)). They have a common reaction mechanism, the metal ion (or ions) activate a water molecule for nucleophilic attack to the substrate (Wilson et al. *Biochemistry* 32, 1659-1694 (1992); Hong & Raushel *Biochemistry* 35, 10904-10911 (1996)); 30 Volbeda et al. *Curr. Opin. Struct. Biol* 6, 804-812; O' Brien & Herschlag *Chemistry & Biology* 6 (1999), and they are all in Class II (Tables II and III).

Changes in residue spacing plays a major role in evolution of protein function, with insertions and deletions contributing substantially to the diversification of enzyme activities. At one level in the α/β -barrel family, such changes can lead to changes in specificity although retaining membership of class I or II. An example is the enolase superfamily (Class II) (Gerlt & Babbitt et al. *Curr Opin Chem Biol* 2, 607-612 (1998); O' Brien & Herschlag *Chemistry & Biology* 6 (1999)'). During evolution, they have retained the structural strategy of catalysing the chemically difficult step of α -proton abstraction but they gained additional functional groups to catalyse different overall reactions (Gerlt & Babbitt et al. *Curr Opin Chem Biol* 2, 607-612 (1998); Gulick et al. *Biochemistry* 37, 14358-14368 (1998)'). Further, more radical, changes can lead to the change of lid design, accompanied by a change in class and a change in mechanism or evolve new function e.g. those with PRAI and IGPS (Hommel et al. *Biochemistry* 34, 5429-5439 (1995); Darimont et al. *Protein Science* 7, 1221-1232 (1998)').

The two classes may be further subdivided on basis of their catalytic mechanism (Table III). Class II barrels, for example, may also be divided into several families, following the criteria used in the SCOP database (Table II) (Murzin et al. *J. Mol. Biol.* 247, 536-540 (1995)'). Some of our class II barrels may be readily subdivided into some of Farber's categories: groups A, D, E and F fit the IGPS group. There is also a correlation between our categories and the description of the β -barrels of Chothia and et al. (Murzin et al. *J. Mol. Biol.* 236, 1369-1381 (1994)) based on packing: our class I corresponds to the distorted TIM

barrel, and the class II encompasses glycolate oxidase and rubisco.

Thus, Nature may have used a three-fold combinatorial
5 strategy for evolving new catalytic activities from pre-
existing α/β -barrel enzymes: retention of mechanism for the
rate determining step but mutation of the binding
specificity (e.g. the formation of the enolate intermediate
in the enolase superfamily Neidhart et al. *Nature* 347, 692-
10 694 (1990) and Neidhart et al. *Biochemical Society Symposia*
57, 135-141 (1990)); retention of binding specificity but
radical mutation of the lid by insertions, deletions and
recombination to change the reaction or its mechanism (e.g.
class I and II aldolases, TIM and Xylose isomerase, PRAI and
15 IGPS Gerlt & Babbitt et al. *Curr Opin Chem Biol* 2, 607-612
(1998) ^{and} O' Brien & Herschlag *Chemistry & Biology* 6 (1999));
and more general changes in the binding site that allow the
catalysis of a variety of different reactions with similar
mechanisms, such as in the superfamily of the metal-
20 dependent hydrolases 'Gerlt & Babbitt et al. *Curr Opin Chem
Biol* 2, 607-612 (1998); Holm & Sander *Proteins* 28, 72-82
(1997)'.

In view of these observations, the inventors now provide
25 practical guidance for the design of new proteins, based on
 α/β -barrels as scaffolds. Once the type of lid Nature uses
for catalysing a particular type of reaction is known, such
a lid can be used as a template for catalysing further
examples of that type of reaction by grafting it onto an
30 α/β -barrel of known binding site. As explained already, this
provides for a general strategy for evolving a new function
in an α/β -barrel scaffold using a combinatorial approach: a
reaction-specific lid is combined with a substrate-specific

binding barrel and subjected to mutation and selection. This approach is particularly well suited for the manipulation of successive enzymes in biosynthetic pathways since the product of one enzyme is the substrate for the next so they 5 both have a common substrate binding site. As described in the following experimental Section 2, this strategy was successfully used to evolve *in vitro* a new function in the α/β -barrel of indole-3-glycerol phosphate synthase and create a novel phosphoribosylanthranilate isomerase of 10 activity comparable to that of the natural enzyme.

SECTION 2

PROVISION OF A NEW ENZYME USING AN α/β -BARREL SCAFFOLD

15 Phosphoribosylanthranilate isomerase (PRAI) activity was evolved from the scaffold of indole-3-glycerol-phosphate synthase (IGPS) by combining a preexisting binding site for structural elements of phosphoribosylanthranilate with a catalytic template required for the isomerase activity. The 20 template was targeted for *in vitro* mutagenesis and recombination, followed by *in vivo* selection. The newly evolved phosphoribosylanthranilate isomerase has similar catalytic properties to the natural enzyme, with an even higher specificity constant.

25 IGPS and PRAI form two covalently linked domains of a bifunctional enzyme in *Escherichia coli* that catalyses two consecutive steps in the tryptophan biosynthesis pathway¹² (Figure 2). The enzymes have a sequence identity of 22% and 30 share a common ligand: carboxyphenylamino-1-deoxy-ribulose 5-P (CdRP), which is the product of PRAI and the substrate of IGPS. There are considerable structural differences between them: IGPS does not isomerise PRA, and PRAI does not

catalyse the formation of the indole ring (Orengo et al. *Structure* 5, 1093-1108 (1997). 14., Holm & Sander *Nucleic Acids Res.* 22, 3600-3609 (1994). 15., Holm & Sander *TIBS* 20, 478-480 (1995)).

5

Design strategy

There are many methods for generating diversity in a target gene (Arnold & Volkov *Curr Opin Chem Biol* 3, 54-59 (1999); Stemmer *Proc. Natl. Acad. Sci. USA* 91, 10747-10751 (1994); 10 Stemmer *Nature* 370, 389-391 (1994); Zhao & Arnold *Nucleic Acids Res.* 25, 1307-1308 (1997); Shao et al. *Nucleic Acids Res.* 26, 681-683 (1998);

Giver & Arnold *Curr Opin Chem Biol.* 2, 335-338 (1998); Zhao et al. *Nat. Biotechnol* 16, 258-261 (1998)). However the

15 generation of mutants must be coupled to a suitable selection procedure for *in vitro* evolution (Arnold & Volkov *Curr Opin Chem Biol* 3, 54-59 (1999); Crameri et al. *Nat. Biotechnol* 14, 315-319 (1996); Crameri et al. *Nat. Medicine* 2, 100-102 (1996); Crameri et al. *Nature* 391, 288-291

20 (1998); Tawfik & Griffiths *Nat. Biotechnol* 16, 652-656 (1998)). A library encoding just one copy of each possible variant for a protein of 250 amino acids (the size of the α/β -barrel) would contain 20^{250} variants, a number constituting a mass far greater than that of known universe

25 (Kauffman, S. A. (ed.) *The origins of order* (Oxford University Press, New York, 1993)). This constraint necessitates both in Nature and in the laboratory the use of techniques that target specially selected segments of the chosen starting scaffold. That is to say a combination of 30 rational design and selection in the experimental strategy for *in vitro* evolution.

The inventors used elements of a pre-existing binding site for the phosphate and anthanilate structural motifs. CdRP

is the product of PRAI, and so the binding site of PRAI must also bind CdRP. IGPS binds CdRP and so the inventors reasoned that it has the potential to bind PRA.

5 The next component of the design was derived from the detailed comparative analysis of structural and biochemical data on IGPS and PRAI by Kirschner and coworkers (Kirschner et al. *Meth. Enzymol.* 142, 386-397 (1987); Darimont et al. *Protein Sci.* 7, 1221-1232 (1998); Hommel et al. *Biochemistry* 10, 34, 5429-5439 (1995); Wilmanns et al. *J. Mol. Biol.* 223, 477-507 (1992); Wilmanns et al. *Biochemistry* 30, 9161-9169 (1991); Knöchel et al. *J. Mol. Biol.* 262, 502-515 (1996); Luger et al. *Science* 243, 206-210 (1989); Stehlin et al. *FEBS Letters* 403, 268-272 (1997)). Using this information, 10 we superimposed the structures of IGPS and PRAI using the program SETOR. From this comparison the active site lid in each protein was identified. The IGPS active site is covered by the N-terminal α 0 helix, and by the β 1- α 1 (15 residues), β 2- α 2 (9 residues) and β 6- α 6 (11 residues) loops, 15 20 all located at the C-terminal side of the β -barrel. PRAI, however, has a very different active site lid which is mainly formed by the β 2- α 2 (10 residues), β 6- α 6 (11 residues) and β 8- α 8 (12 residues) loops. The β 2- α 2 loop is involved in 25 binding the anthranilic acid moiety of the substrates PRA and CdRP, and the β 8- α 8 loop comprises the phosphate binding site. The superposition of the two structures reveals almost identical locations but different orientations of the phosphate binding site. Since the loops (β 2- α 2, β 7- α 7 and β 8- α 8) are similarly arranged in the two enzymes, the target of 30 selection was solely the extra N-terminal end (helix α 0 and two bends), the β 1- α 1 loops and the β 6- α 6 loops.

The first step in the design included the deletion of 48 amino acid residues from the amino terminal end of IGPS; this deletion mutant was called (IGPS49). This mutant was unstable, had a tendency to aggregate (Stehlin et al. *FEBS Letters* 403, 268-272 (1997)) and was catalytically inactive with respect to both IGPS and PRAI activities. Nucleic acid encoding the IGPS49 was expressed in *E. coli*, and the protein formed inclusion bodies. Refolding chromatography with immobilised minichaperones was employed to renature the protein quantitatively (Altamirano et al. *Proc. Natl. Acad. Sci. USA* 94, 3576-3578 (1997)). It had a circular dichroism (CD) spectrum characteristic of a native α/β -barrel protein and bound ^3H -rCdRP, a specific inhibitor of IGPS, with a stoichiometry of one mol of inhibitor per mol of IGPS49.

15

The IGPS49 scaffold was further modified by replacing 15 amino acid residues corresponding to the β 1- α 1 loop by a new randomised segments of 4 to 7 amino acid residues. Nucleic acid encoding IGPS49 was used as template to create three new libraries IGPS49L1 (GKXXG), IGPS49L1RGD (GKXRGD) and IGPS49L1SV (length size variation: GKXX, GKXXX, GKXXXX or GKXXXXX) via PCR methodologies including overlap extension PCR, inverse PCR and random primer PCR. The libraries were analysed by PCR screening, by restriction analysis and by sequencing. Members of each library were picked at random and expressed in *E. coli*. The proteins appeared in the soluble fraction but were prone to aggregation above a concentration of 0.5 mg/mL. One of the protein samples was denatured in 8 M urea and renatured using refolding chromatography (Altamirano et al. *Proc. Natl. Acad. Sci. USA* 94, 3576-3578 (1997)). The refolded protein was soluble and able to bind ^3H -rCdRP, but it lacked catalytic activity.

The next set of modifications involved the β 6 α 6 loop, including the introduction of an aspartic residue at position 184 (acting as a general base in the active site) (Darimont et al. *Protein Sci.* 7, 1221-1232 (1998); Wilmanns 5 et al. *J. Mol. Biol.* 223, 477-507 (1992)) and also the PRAI consensus sequence GXGGXGQ (Wilmanns et al. *J. Mol. Biol.* 223, 477-507 (1992)), with the aim of improving the active site lid. A new library including these modifications and called IGPS49L1L6 was constructed using the IGPS49L1, 10 IGPS49L1RGD and IGPS49L1SV libraries as templates. One of the new library members chosen at random was expressed in *E. coli* and the corresponding protein was found to be soluble, with a circular dichroism spectrum characteristic of a typical α / β -barrel protein. Further, it was able to bind the 15 3 H-rCdRP, but lacked either PRAI or IGPS activity.

Mutation, recombination and in vivo selection

An *in vivo* selection strategy for PRAI activity was designed, based on complementation of *E. coli* JA300 (a PRAI-20 deficient strain that does not grow in the absence of tryptophan (Trp), and which is available from ATCC). In *E. coli*, PRAI and IGPS are part of the same 45 kDa polypeptide chain specified by the *trpC* gene. However, *E. coli* JA300 carries the W3110(*trpC1117*) allele and so lacks isomerase 25 activity, but retains normal levels of synthase activity (Clarke *Proc. Natl. Acad. Sci. USA* 77, 2173-2177 (1980); Yanofsky et al. *Genetics* 69, 409-433 (1971); Yanofsky *JAMA* 218, 1026-1035 (1971)). Complementation provides indication 30 that the specific clone contains a plasmid expressing an IGPS variant with PRAI activity.

JA300 itself, showed no ability to grow in the absence of Trp. The initial parental clones (IGPS49, IGPS49L1,

IGPS49L1RGD, and IGPS49LSV) failed to grow in absence of Trp.

The DNA library IGPS49L1L6 was used to transform the JA300 strain. Approximately 3×10^4 *E. coli* transformants expressing the resultant library were then plated on minimal medium containing a range of tryptophan concentrations (0-25 $\mu\text{g/mL}$). The colonies (around 500) growing at low Trp concentrations were selected. A first round of DNA shuffling was performed with the pool of genes from the selected clones using the method of Stemmer (Stemmer *Proc. Natl. Acad. Sci. USA* 91, 10747-10751 (1994); Stemmer *Nature* 370, 389-391 (1994); Crameri et al. *Nat. Biotechnol* 14, 315-319 (1996)). Plating around 4×10^5 bacteria on a wide range of Trp concentrations yielded 80 colonies. These were able to grow at very low concentration of Trp ($< 1 \mu\text{g/mL}$) and a single clone was found to be capable of growing in the absence of any exogenous Trp. Restriction-fragment length polymorphism (RFLP) analysis of 30 clones chosen at random revealed a minimum of 8 different patterns. A second round of recombination was performed by DNA shuffling (Stemmer *Nature* 370, 389-391 (1994)) and staggered extension procedure (StEP) (Zhao et al. *Nat. Biotechnol* 16, 258-261 (1998)), using the pool of 80 colonies selected from the first round and synthetic DNA fragments encoding for the protein segments corresponding to loops $\beta 1\alpha 1$, $\beta 6\alpha 6$, $\beta 4\alpha 4$ from diverse species of PRAI. The *in vivo* selection yielded 360 colonies capable of growing in the absence of any exogenous Trp.

30

Several controls were performed in order to show that the ability to grow in absence of Trp was a consequence of the

introduction of the reshuffled library (IGPS49L1L6-2 cycle) containing the *ivePRAI* (*in vitro*-evolved PRAI) genes. As a first control, the inventors cured the JA300 strain previously transformed with the plasmids carrying the 5 library by growing the bacteria in the absence of ampicillin. The cured cells were unable to grow in ampicillin-containing medium and simultaneously lost the ability to grow in absence of Trp. Further, the plasmid carrying the *ivePRAI* gene was used to transform fresh JA300 10 cells, prior to plating on minimal medium with added ampicillin, Streptomycin (Strep) and IPTG but in the absence of Trp. These transformed cells were able to grow in 18 h in the absence of Trp, all the clones were ampicillin resistant and were Trp⁺ (see additional controls in the 15 Materials and Methods section below). On the basis of these controls, it is believed that the PRAI activity complementing the auxotrophy in JA300 cells originates from the cloned IGPS variant genes and is not the product of any reversion event.

20

In vitro-evolved PRAI

The nucleic acid encoding the *ivePRAI* proteins from 30 clones were sequenced. Only 8 different sequences were found. The largest colony from a plate of minimal medium 25 without Trp was selected for further biochemical characterisation. The gene encoding the *ivePRAI* was expressed and the protein purified. The new protein was soluble. The CD spectra and the activity assay confirmed that was properly folded.

30

The *ivePRAI* has PRAI activity and does not have IGPS activity *in vitro*. *ivePRAI* has a specificity constant (k_{cat}/K_m) of $4.8 \times 10^7 \text{ s}^{-1} \text{ M}^{-1}$ (Table I), which is 6-fold higher

than that of either the natural enzyme (*E. coli* wild-type bienzyme) or the isolated PRAI domain (Table I). This improved activity results primarily from a 15-fold enhanced affinity of the evolved protein for PRA (Table I).

5

The structure of *ivePRAI* resembles IGPS and differs significantly from that of PRAI. The sequence identity for *ivePRAI* to PRAI is 28% and is 90% to IGPS (Figure 3). Importantly, the binding site for the phosphate ion in the 10 IGPS scaffold of *ivePRAI* is at the N-terminal turn of the additional α -helix $\alpha 8'$ that is located in the loop between strand $\beta 8$ and helix $\alpha 8$. In the wild-type PRAI, the additional α -helix $\alpha 8'$ is missing and the phosphate ion has different orientation (Wilmanns et al. *J. Mol. Biol.* 223, 15 477-507 (1992)). Further, the site for binding the anthranilate moiety of PRA in *ivePRAI* is also inherited from IGPS and is quite different from that of PRAI. The catalytic constants of *ivePRAI* and PRAI are similar (Table 1).

20 These experiments demonstrate that the two classes of α/β -barrels, described above, can be interconverted by altering the lid regions. The results demonstrate the divergent evolution of two enzymes from the pathway for the biosynthesis of tryptophan, which may mimic natural 25 divergent evolution (Sterner et al. *Protein Science* 5, 2000-2008 (1996)). For *in vitro* design purposes, a new function in the scaffold of an α/β -barrel protein was provided using the combined approach of rational design, *in vitro* mutation, recombination and *in vivo* selection.

30

MATERIALS AND METHODS

Reagents

Restriction enzymes and T4 DNA ligase were obtained from BioLabs. Taq polymerase and Wizard DNA preparation kits were obtained from Promega. Ultrapure dNTPS were obtained from 5 Boehringer Mannheim. DNase I and other reagents were obtained from Sigma.

Chemical syntheses

10 rCdRP and 3 H-rCdRP were prepared as described by Bisswanger et al. (Bisswanger et al. *Biochemistry* 18, 5946-5953 (1979)). The specific activity of the 3 H-rCdRP was 95.36 kBq/ μ mol.

Preparation of DNA

15 The gene encoding IGPS (residues 1-259) was amplified from *E. coli* BL21 genomic DNA by PCR (94 °C, 1 min; 37 °C, 1 min; 72 °C, 1 min; 25 cycles) using primers 'IGPSFULL' and 'IGPSFLAGREV'. The PCR product was digested with *Nco* I and *Bsp* HI and the 820bp fragment cloned in to the *Nco* I site of 20 pNS3785 (Sternberg et al. *Proc. Natl. Acad. Sci. USA.* 92, 1609-1613 (1995)) to create pJB122. pJB122 thus encodes a polypeptide chain comprised of residues 49-259 of IGPS fused directly to the Flag-tag GSDYKDDDDK at the C-terminus of IGPS. The gene encoding IGPS49 (residues 49-259) was 25 amplified by PCR from pJB122 using primers 'IGPS49FSP1' and 'JB122SEQ' and was then digested with *Fsp* I and *Bam* HI. pJB124 was created by ligation of the 630bp PCR fragment with a 4700bp fragment generated from pJB122 by digestion with *Nco* I, blunt-ending with Klenow polymerase, and further 30 digestion with *Bam* HI. The gene encoding IGPS49 was used as a template for further modifications and recloned in the same vector described above, a set of different plasmids (pMA) carrying all the libraries were created.

Oligonucleotides

The following oligonucleotides were used.

5 IGPSSFULL:

5'CATGACCTTGCAGCCAGCCGGCATGGCGAAACGTTTAGCGAAAATCGTCGC3'

IGPSFLAGREV:

5' ATCGTCATAATCATGAACACTTGTATCGTCGTCCTGTAGTCGGATCCTACTTTAT
TCTCACCCAGCAACACCCGGCGCACGG3'

IGPS49L1:

5'NNNSNNNSGGTGCACGCATTGCCGCCATTATAAACATTACGC3'

15 IGPS49Lr:

5'ACCGCACTCCAGAATAATGCCCTTCC3'

IGPS49FSP1:

5' CATGACCTTGTGCGCATTATTCTGGAGTGC3'

20

JB122SEQ:

5 'CCCTGCGGCTGGTAATGG3 '

IGPS49L1L6r:

25 5' CCCACCSNNGCCGTTGATGCCAACGACCTTGCCCC3'

IGPS (Apali) :

5 'CGCCGTGCGTGCACCTGTAGCGC3 '

30 L1(6aa):

5' GGAAGGGCATTTATTCTGGAGTGCAGTNNNSNNSGGTGCACGCATTGCCGCC3'

L1APAL1:

5' TTTATTCTGGAGTGCCTCTANNNSNNNSGGTGCACGCATTGCCGCC3'

L1APALre:

5'GGCGGCAATGCGTGCACCSNNNSNNNTAGACCGCACTCCAGAATAAA3'

5

L6:

5'GCAAAGGTCGTTGGCATCACGGCNNSGGTGGGNNSGGTNNNSNNSATGATCTCAACC
GTACC3'

10 L6rev:

5'GGTACGGTTGAGATCAATSNNNNACCSNNCCACCSNNNGCCGTTGATGCCAACGACC
TTTGC3'

DNA shuffling

15 The shuffling of the pool of genes from the first cycle of selection was performed using 60 to 80 bp fragments, generated by DNase I (Sigma) and reassembled by PCR without added primers (Stemmer *Proc. Natl. Acad. Sci. USA* 91, 10747-10751 (1994)). A PCR program of 95 °C, 1 min, 40 cycles (94 °C, 30 s; 55 °C, 30 s; 72 °C, 1 min + 5 sec. per cycle) was used. After 40-fold dilution of the minus primer product into PCR mix with 1 µM of each primer and 20 additional cycles of PCR (94 °C, 30 s; 55 °C, 30 s; 72 °C, 2 min), a single product of 650 bp was obtained. The shuffled material 20 was cloned back into the vector described above and used to transform the PRAI-deficient *E. coli* strain JA300 (Clarke *Proc. Natl. Acad. Sci. USA* 77, 2173-2177 (1980); Yanofsky & Horn *J. Bacteriol.* 176, 6245-6254 (1994)).

25 The second cycle of shuffling was performed on the pool of chimaera selected in the first round and synthetic DNA fragments encoding for the protein segments corresponding to loops β1α1, β6α6, β4α4 from diverse species of PRAI.

Staggered extension process (StEP)

The StEP conditions were performed as described in Zhao et al. *Nat. Biotechnol* 16, 258-261 (1998). A PCR program of 92 cycles (94 °C, 30 sec; 55 °C, 4 sec) was used. At this step the parent DNA (purified from a *dam*⁺ strain) was removed using *Dpn* I. A second PCR was performed adding primers in order to amplify the full length product (95 °C, 2 min; 25 cycles (94 °C, 30 sec; 55 °C, 1 min; 72 °C, 5 min) 72 °C, 30 min).

Selection

JA300 cells were plated on minimal medium (M9) with ampicillin (50 μ g/mL), streptomycin (20 μ g/mL) plus 0.7 mM IPTG, containing a range of Trp concentration and incubated at 37 °C for 24-36 h. About 500 colonies from the plates with the lower Trp levels were pooled and cultured either in liquid medium 2X TY + amp + Strep or minimal medium (M9) + Amp + Strep + 0.7 mM IPTG with the similar level of Trp. Plasmid DNA was prepared from this liquid culture.

Additional controls experiments:

Plasmid DNA from the pool of clones selected after the second round of recombination was prepared and used DNA to transform fresh JA300 cells, prior to plating on minimal medium with added ampicillin, streptomycin (Strep) and IPTG but in the absence of Trp. These transformed cell were able to grow in the absence of Trp in 18 h. Additionally, the plasmid DNA from these cells was purified and the insert excised by restriction digestion and recloned into a fresh vector. After transforming into fresh JA300 cells, positive clones were obtained in the absence of Trp, demonstrating that the activity was insert dependent. The same result was

obtained when the DNA was amplified by PCR, recloned and introduced into fresh JA300 cells.

Refolding chromatography

5 Protein renaturation was performed as described in Altamirano et al. *Proc. Natl. Acad. Sci. USA* 94, 3576-3578 (1997).

Protein purification

10 After refolding experiments, the proteins (IGPS49, IGPS49L1 and IGPS49L1L6) were purified as described in Bisswanger et al. *Biochemistry* 18, 5946-5953 (1979).

PRAI activity Assay

15 All the kinetic and binding experiments were performed as described in Kirschner et al. *Meth. Enzymol.* 142, 386-397 (1987) and Hommel et al. *Biochemistry* 34, 5429-5439 (1995).

Sequence alignment

20 The amino acid sequences of *ivePRAI*, IGPS and PRAI were aligned using sequence similarity search of SCOP sequences based on BLAST algorithm Stephen et al. *J. Mol. Biol.* 215, 403-410 (1990). In Figure 3 we show the sequence alignment based on ClustalW algorithm (Matrix Blosum 30).

TABLE I CLASS I α/β -BARREL PROTEINS - NO. OF RESIDUES IN DIFFERENT REGIONS

PROTEINS	amino terminal extension	Connection between β_1 - α_1	Connection between β_2 - α_2	C-terminal after β_8
phosphoribosyl anthranilate isomerase (PRAI)	5	7	11	19
Triose Phosphate isomerase (TIM)	6	6	3	17
Phosphoenolpyruvate /pyruvate domain:	3	2	5	24
Pyruvate kinase M1 Pyruvate kinase	5	7	5	24
Class II aldolase Fructose-biphosphate aldolase Luciferase	9	3	19	92
Flavoprotein 390	5	11	22	32

TABLE II CLASS II α/β -BARREL PROTEINS - NO. OF RESIDUES IN DIFFERENT REGIONS

PROTEINS	amino terminal extension	Connection between $\beta_1-\alpha_1$	Connection between $\beta_2-\alpha_2$	C-terminal after β_8
<i>Tryptophan biosynthesis enzymes</i>				
IGPS	48	14	12	21
α -subunit of tryptophan synthase	14	8	26	39
<i>Glycosyltransferase</i>				
Endo-1,4-beta-d-glucanase	39	4	18	29
Alpha-amylase, high pi isozyme	2	21	27	24
Narbonin	3	20	18	22
Xylose isomerase	10	22	10	102
<i>NADP-linked oxidoreductase</i>				
Inosine monophosphate dehydrogenase	53	8	5	100
<i>Class I aldolases</i>				
Fructose 1,6 biphosphate aldolase	30	5	8	70
N-acetylneuraminate lyase	4	14	12	83

Metal-dependent hydrolases	51	19	10	64
Phosphotriesterase				
Aldose reductase	17	6	7	55
Adenosine deaminase	10	47	24	61
Dihydroorotate	68	11	10	131
Methylmalonyl-CoA mutase				
(Chain A)	84	5	29	50
tRNA-guanine				
transglycosylase	44	10	10	101
FMN-linked oxidoreductase				
Glycolate oxidase	72	12	19	51
Old yellow enzyme	31	20	16	54
Trimethylamine				
dehydrogenase	23	12	12	98
Rubisco				
Enolase family				
Yeast enolase	17	12	6	40
Mandelate racemase	20	11	3	50
D-Glucarate Dehydratase	18	22	6	41
Phosphatidylinositol-				
specific phospholipase C	45	22	8	23
(Pi-PLC)				

Table III Correlation between the structural class of the lid and the reaction mechanism

Enzyme	EC number	Reaction mechanism	Lid class
Class II Aldolase Fructose biphosphate aldolase	4.1.2.13	Aldol condensation Metal activation of a carbonyl group.	I
Class I Aldolase Fructose-1-6-biphosphate aldolase	4.1.2.13	Aldol condensation Lysine-shiff base and α -carbon activation.	II
Triosephosphate isomerase	5.3.1.1.	Aldol-ketol isomerase (intramolecular al oxidoreductase) Proton abstraction and 1,2 transfer via an enol intermediate.	I
Xylose isomerase	5.3.1.5	Aldol-ketol isomerase (intramolecular al oxidoreductase, direct 1, 2 hydride transfer).	II
Phosphoribosylanthranilate isomerase	5.3.1.24	Isomerase (intramolecular oxidoreductase) Amadori rearrangement.	I
Indole-3-glycerol phosphate synthase	4.1.1.48	Electrophilic attack Enolisation, descarboxylation and carbanion addition to a double bond. Proton abstraction,	II
Enolase superfamily			II

		water elimination and C=C bond formation.	
<i>Enolase</i>	4.2.1.1.1	"	III
<i>(D)-glucarate</i>	4.2.1.4.0	"	II
<i>Dehydratase</i>			
<i>Muconate lactonizing</i>	5.5.1.1	"	III
<i>enzyme</i>			
<i>Mandelate racemase</i>	5.1.2.2	"	II
<i>Hydrolase superfamily</i>		Metal-bound hydroxide ion.	II
<i>Dihydroorotate</i>	3.5.2.5.	"	II
<i>Adenosine deaminase</i>	3.5.4.4	"	II
<i>Phosphotriesterase</i>	3.5.-.-	"	II

Table IV

<u>2,5-DIKETO-D-GLUCONIC ACID REDUCTASE A</u>
<u>ALCOHOL DEHYDROGENASE (NADP+)</u>
<u>ALDEHYDE REDUCTASE</u>
<u>3-ALPHA-HYDROXYSTEROID DEHYDROGENASE (B-SPECIFIC)</u>
<u>IMP DEHYDROGENASE</u>
<u>3-ALPHA-HYDROXYSTEROID DEHYDROGENASE (A-SPECIFIC)</u>
<u>L-LACTATE DEHYDROGENASE (CYTOCHROME)</u>
<u>(S)-2-HYDROXY-ACID OXIDASE</u>
<u>DIHYDROORotate OXIDASE</u>
<u>TRIMETHYLAMINE DEHYDROGENASE</u>
<u>NADPH DEHYDROGENASE</u>
<u>5,10-METHYLENETETRAHYDROFOLATE REDUCTASE (FADH)</u>
<u>ALKANAL MONOOXYGENASE (FMN-LINKED)</u>
 <u>TRANSALDOLASE</u>
 <u>CYCLOMALTODEXTRIN GLUCANOTRANSFERASE</u>
 <u>NICOTINATE-NUCLEOTIDE PYROPHOSPHORYLASE (CARBOXYLATING)</u>
 <u>QUEUINE tRNA-RIBOSYLTRANSFERASE</u>
 <u>THIAMIN-PHOSPHATE PYROPHOSPHORYLASE</u>
 <u>DIHYDROPTEROATE SYNTHASE</u>
 <u>PYRUVATE KINASE</u>
 <u>PYRUVATE PHOSPHATE DIKINASE</u>

Table IV (Continued)

<u>1-PHOSPHATIDYLINOSITOL PHOSPHODIESTERASE</u>
<u>1-PHOSPHATIDYLINOSITOL-4,5-BISPHOSPHATE PHOSPHODIESTERASE</u>
<u>ARYLDIALKYLPHOSPHATASE</u>
<u>DEOXYRIBONUCLEASE IV (PHAGE T4-INDUCED)</u>
<u>ALPHA-AMYLASE</u>
<u>BETA-AMYLASE</u>
<u>CELLULASE</u>
<u>ENDO-1,4-BETA-XYLANASE</u>
<u>OLIGO-1,6-GLUCOSIDASE</u>
<u>CHITINASE</u>
<u>BETA-GLUCOSIDASE</u>
<u>BETA-GALACTOSIDASE</u>
<u>BETA-GLUCURONIDASE</u>
<u>GLUCAN ENDO-1,3-BETA-D-GLUCOSIDASE</u>
<u>BETA-N-ACETYLHEXOSAMINIDASE</u>
<u>GLUCAN 1,4-ALPHA-MALTOTETRAHYDROLASE</u>
<u>ISOAMYLASE</u>
<u>LICHENINASE</u>
<u>MANNAN ENDO-1,4-BETA-MANNOSIDASE</u>
<u>6-PHOSPHO-BETA-GALACTOSIDASE</u>
<u>CELLULOSE 1,4-BETA-CELLOBIOSIDASE</u>
<u>MANNOSYL-GLYCOPROTEIN ENDO-BETA-N-ACETYLGLUCOSAMIDASE</u>
<u>NEOPULLULANASE</u>
<u>THIOGLUCOSIDASE</u>
<u>UREASE</u>
<u>ADENOSINE DEAMINASE</u>

Table IV (Continued)

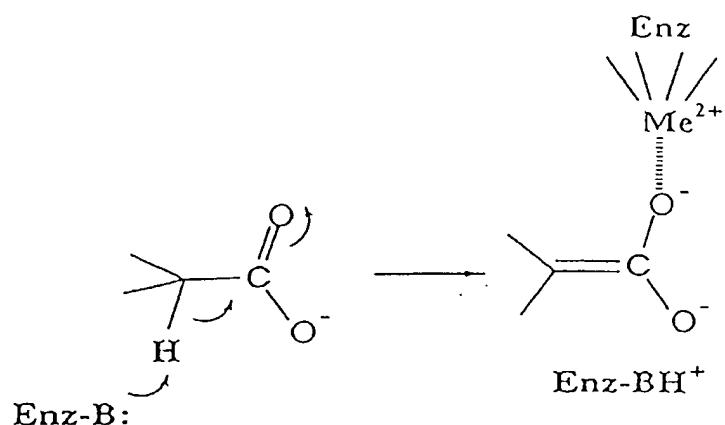
<u>PHOSPHOENOLPYRUVATE CARBOXYLASE</u>
<u>RIBULOSE-BISPOLYPHOSPHATE CARBOXYLASE</u>
<u>INDOLE-3-GLYCEROL-PHOSPHATE SYNTHASE</u>
<u>FRUCTOSE-BISPOLYPHOSPHATE ALDOLASE</u>
<u>2-DEHYDRO-3-DEOXYPHOSPHOGLUCONATE ALDOLASE</u>
<u>2-DEHYDRO-3-DEOXYPHOSPHOHEPTONATE ALDOLASE</u>
<u>N-ACETYLNEURAMINATE LYASE</u>
<u>PHOSPHOPYRUVATE HYDRATASE</u>
<u>TRYPTOPHAN SYNTHASE</u>
<u>PORPHOBILINOPENICYLIC ACID SYNTHASE</u>
<u>GLUCARATE DEHYDRATASE</u>
<u>DIHYDRODIPICOLINATE SYNTHASE</u>
<u>ALANINE RACEMASE</u>
<u>MANDELATE RACEMASE</u>
<u>RIBULOSE-PHOSPHATE 3-EPIMERASE</u>
<u>TRIOSEPHOSPHATE ISOMERASE</u>
<u>XYLOSE ISOMERASE</u>
<u>PHOSPHORIBOSYLANTHRANILATE ISOMERASE</u>
<u>PHOSPHOENOLPYRUVATE MUTASE</u>
<u>METHYLMALONYL-COA MUTASE</u>
<u>MUCONATE CYCLOISOMERASE</u>
<u>CHLOROMUCONATE CYCLOISOMERASE</u>

Table IV (Continued)

<u>CONCANAVALIN B</u>
<u>NARBONIN</u>
<u>NONFLUORESCENT FLAVOPROTEIN</u>
<u>PHOSPHOTRIESTERASE HOMOLOGY PROTEIN</u>
<u>YEAST HYPOTHETICAL PROTEIN</u>

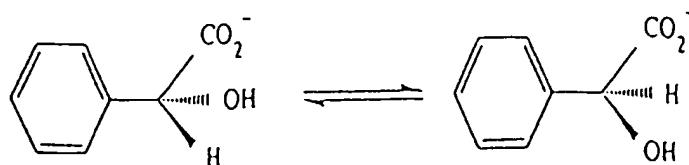
Annex 1

Nature strategies to evolve new proteins. Nature may have used a combinatorial strategy for evolving new catalytic activities from pre-existing α/β -barrel enzymes. These strategies, are at least three:



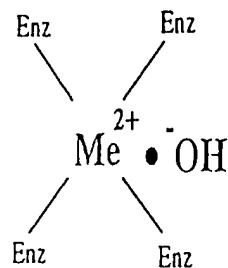
1) A rate-limiting step in the catalytic mechanism is retained and the substrate-binding site (the hydrophobic pocket and charged region) evolved by punctual mutations. For instance, in the enolase superfamily. 16,17

The fate of the intermediate is determined by the structure of each active site, so that the overall reactions differ and may involve 1,1-proton transfer (racemization): Mandelate racemase

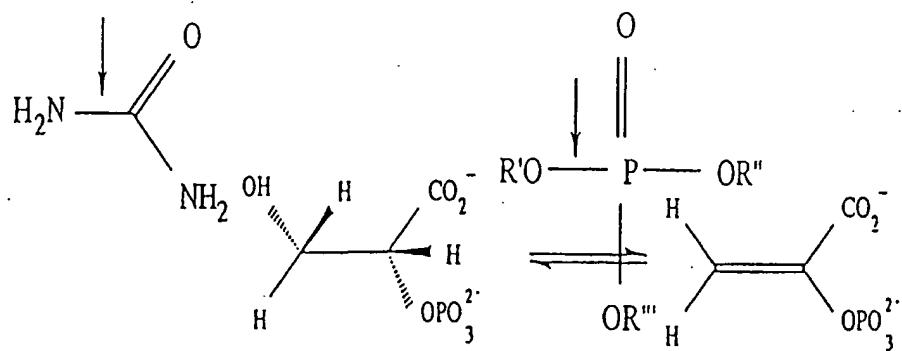


or β -elimination of water: Enolase

2) In the superfamily of the metal-depend hydrolases the general mechanistic features are conserved (e.g. metal binding site)^{7,8} but few changes in the charged and hydrophobic regions of the binding site allows the catalysis of multitude of different reactions.



Overall reactions:



Urease

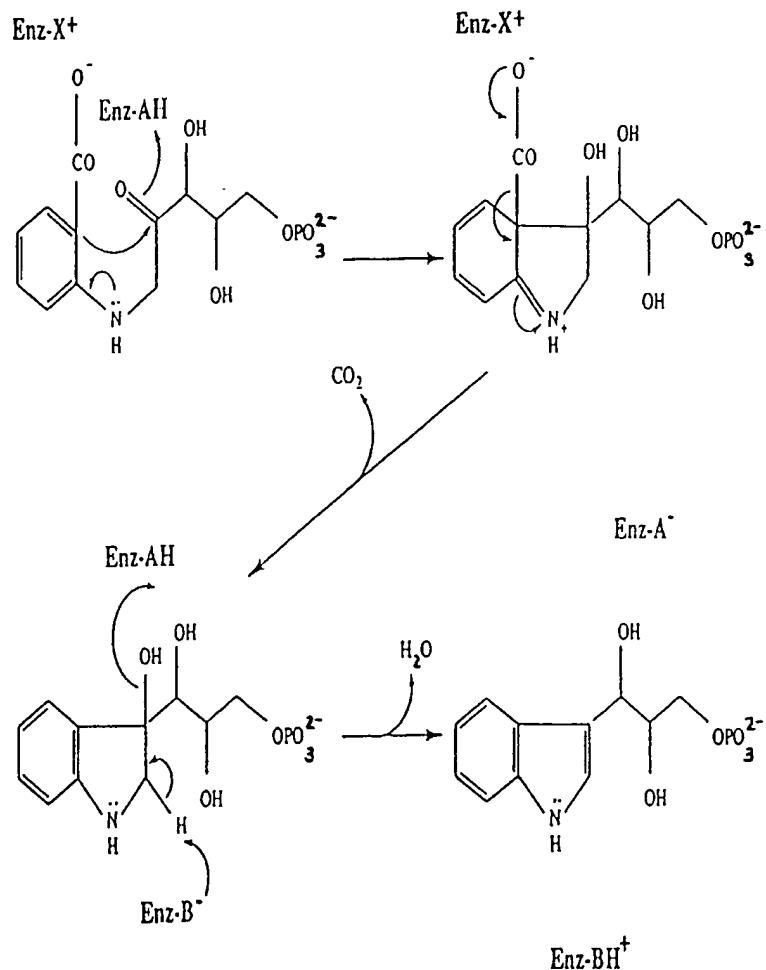
Phosphotriesterase

73

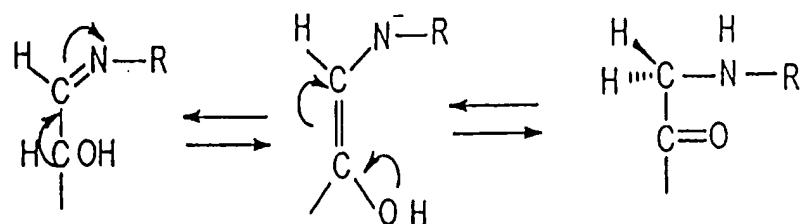
3). In Class I Aldolases and Class II aldolases, TIM and Xylose isomerase, PRAI and IGPS.^{7,12}

the structure of the binding site may be retained and that of the active-site lid is modified by insertions, deletions and recombination.

IGPS:



PRAI:



CLAIMS

1. A method of obtaining an enzyme that catalyses a desired reaction on a target substrate, the method comprising:

5 selecting a parent α/β barrel enzyme that comprises a scaffold and an active site lid and which either
 (i) binds the target substrate, or
 (ii) binds a similar substrate and catalyses a reaction
 of the same type as said desired reaction;

10 modifying the amino acid sequence of the N-terminal segment, β_1 - α_1 loop, β_6 - α_6 loop and/or C-terminal segment of the parent α/β barrel enzyme, and optionally altering additional amino acid residues within the parent α/β barrel enzyme, whereby one or more candidate product enzymes is
15 obtained;

 selecting from the candidate product enzymes a product enzyme that comprises a scaffold and an active site lid, which product enzyme catalyses the desired reaction on the target substrate.

20 2. A method according to claim 1 wherein the parent enzyme comprises a scaffold that binds the target substrate.

25 3. A method according to claim 1 or claim 2 wherein said modifying of the parent enzyme to obtain one or more candidate product enzymes comprises grafting to the scaffold of the parent enzyme an active site lid of another enzyme.

30 4. A method according to any one of claims 1 to 3 comprising modifying the parent α/β barrel enzyme by deleting an N-terminal segment, shortening the β_1 - α_1 loop, and modifying the β_6 - α_6 loop.

5. A method according to any one of claims 1 to 3 comprising modifying the parent α/β barrel enzyme by adding an N-terminal segment, lengthening the $\beta 1-\alpha 1$ loop, and modifying the $\beta 6\alpha 6$ loop.

5

6. A method according to any one of claims 1 to 5 comprising modifying an N-terminal segment, the $\beta 1-\alpha 1$ loop, and the $\beta 6\alpha 6$ loop, and optionally altering one or more amino acid residues within one or more of the loops $\beta 3-\alpha 7$, $\beta 7-\alpha 7$ and $\beta 5-\alpha 5$.

10

7. A method according to any one of claims 1 to 5 comprising altering one or more amino acid residues between the loops $\beta 7-\alpha 7$ and $\beta 8-\alpha 8$.

15

8. A method according to any one of claims 1 to 5 comprising altering one or more amino acid residues in one or more of the loops $\beta 2-\alpha 2$, $\beta 4-\alpha 4$ and $\beta 3-\alpha 3$.

20

9. A method according to any one of claims 1 to 8 comprising modifying the parent α/β barrel enzyme to introduce one or more amino acid sequence motifs or residues in accordance with a consensus for α/β barrel enzymes that catalyse the desired reaction or a reaction of the same type as the desired reaction.

25

10. A method according to any one of claims 1 to 9 comprising random mutagenesis of residues within the parent α/β barrel enzyme, and selection of a candidate enzyme on ability to bind said target substrate.

30

11. A method according to any one of claims 1 to 9 comprising random mutagenesis of residues within the parent α/β barrel

enzyme, and selection of product enzyme on ability to catalyse the desired reaction on said target substrate.

12. A method according to any one of claims 1 to 11 further 5 comprising, following the obtaining of said product enzyme, providing nucleic acid encoding the product enzyme.

13. A method according to claim 12 wherein said nucleic acid 10 is provided operably linked to regulatory sequences within an expression vector for expression of the encoded product enzyme.

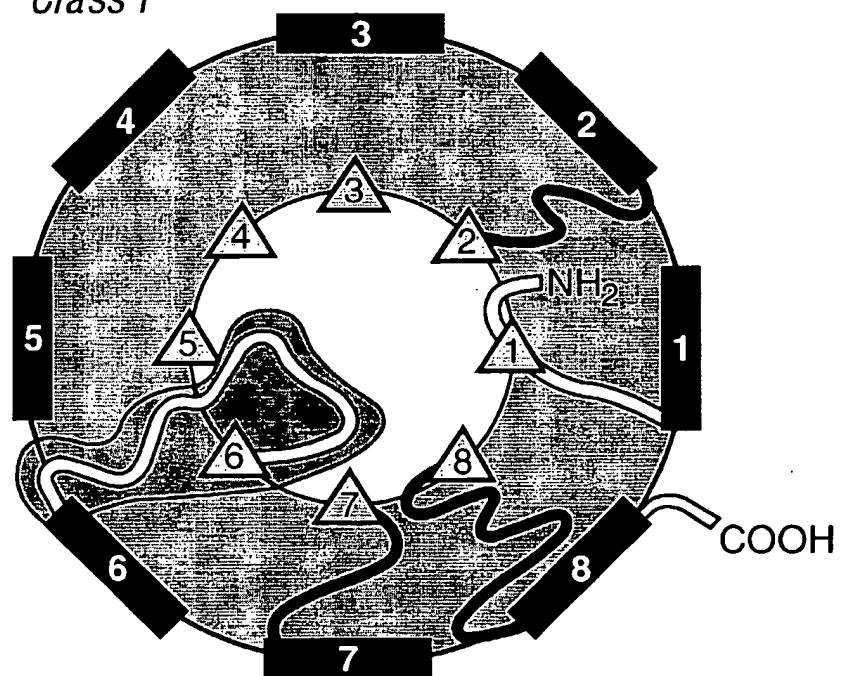
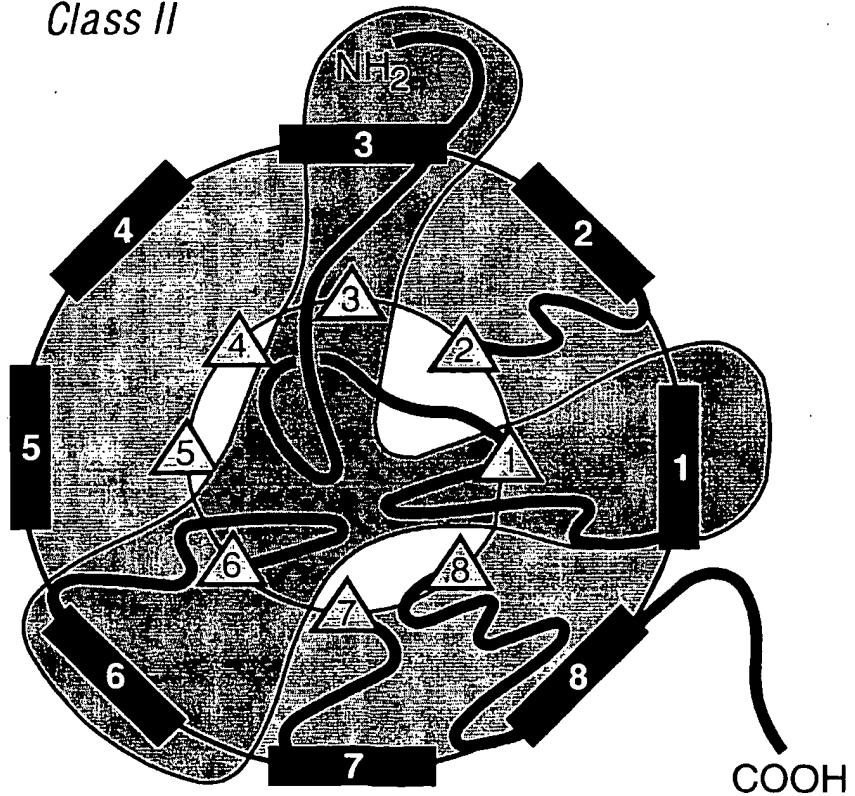
14. A method according to any one of claims 1 to 11 further comprising, following the obtaining of said product enzyme, 15 synthesizing said product enzyme by expression from encoding nucleic acid in a recombinant system.

15. A method according to claim 14 further comprising isolating and/or purifying said product enzyme.

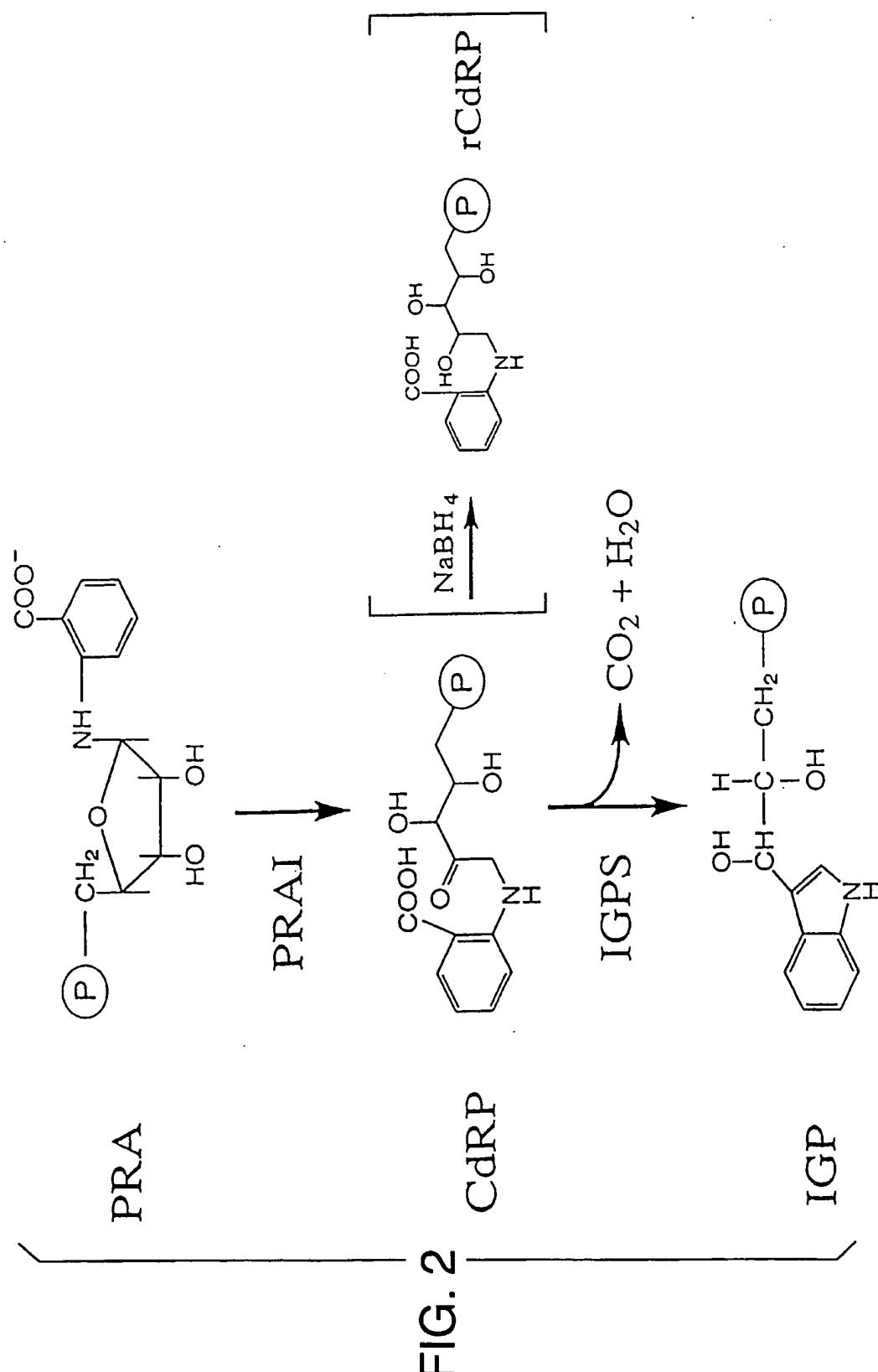
20 16. A method according to claim 14 or claim 15 further comprising formulating said product enzyme into a composition comprising at least one additional component.

1/6

FIG. 1

Class I*Class II*

2/6



4/6

FIG. 4

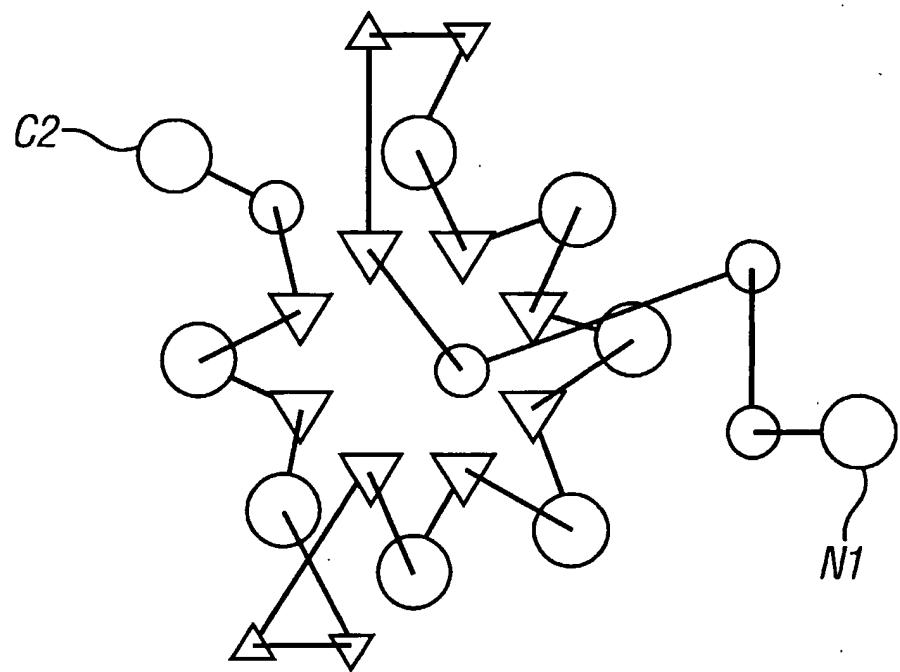


FIG. 5

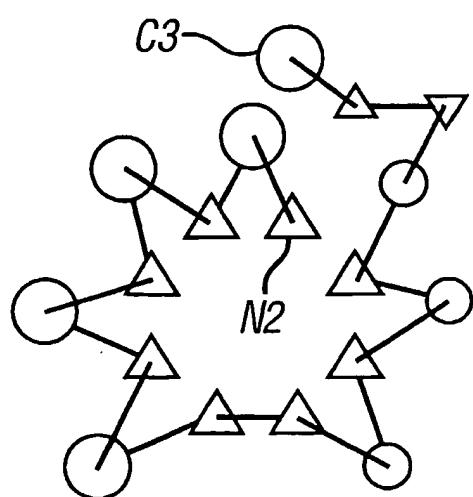
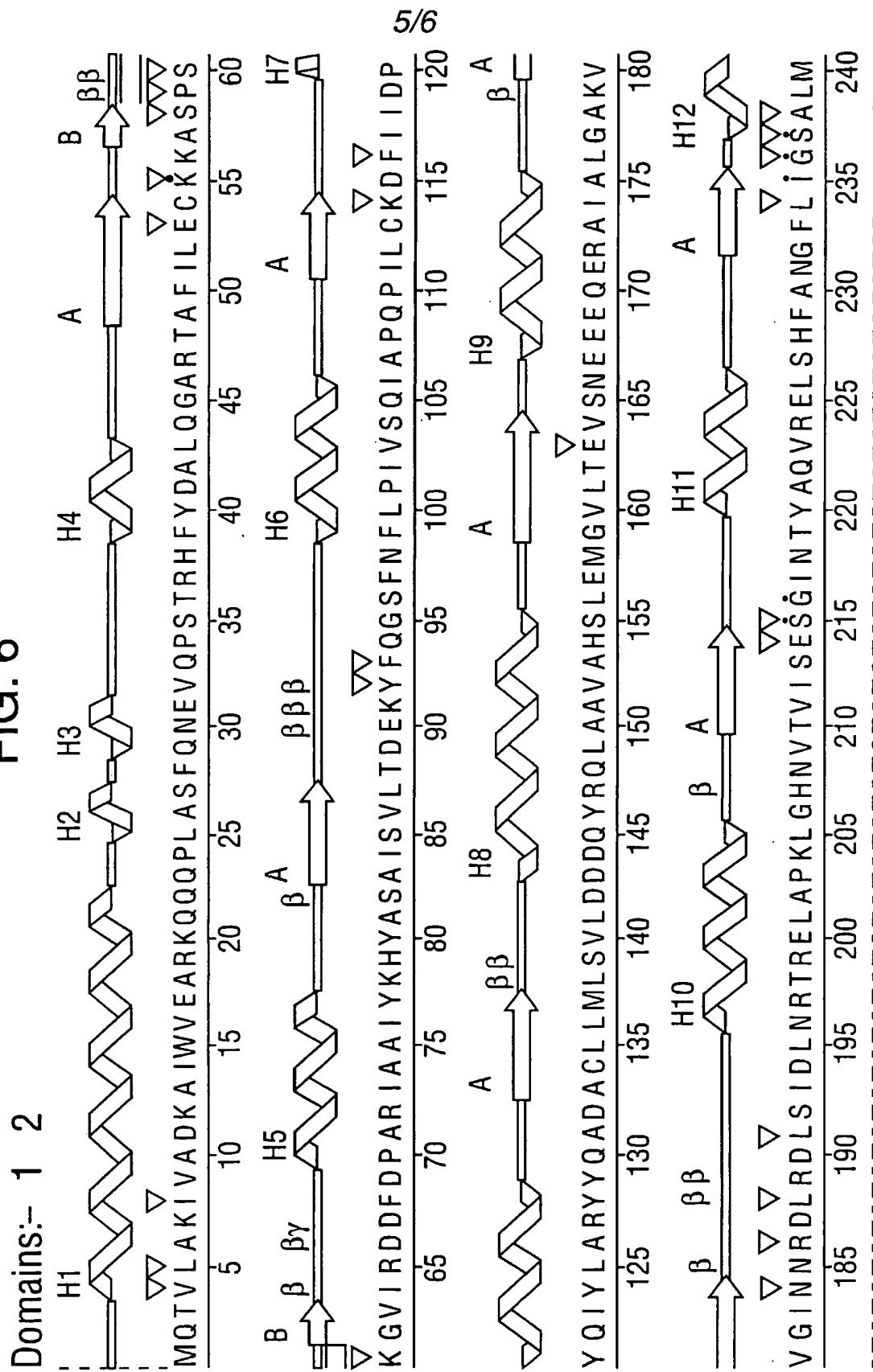


FIG. 6



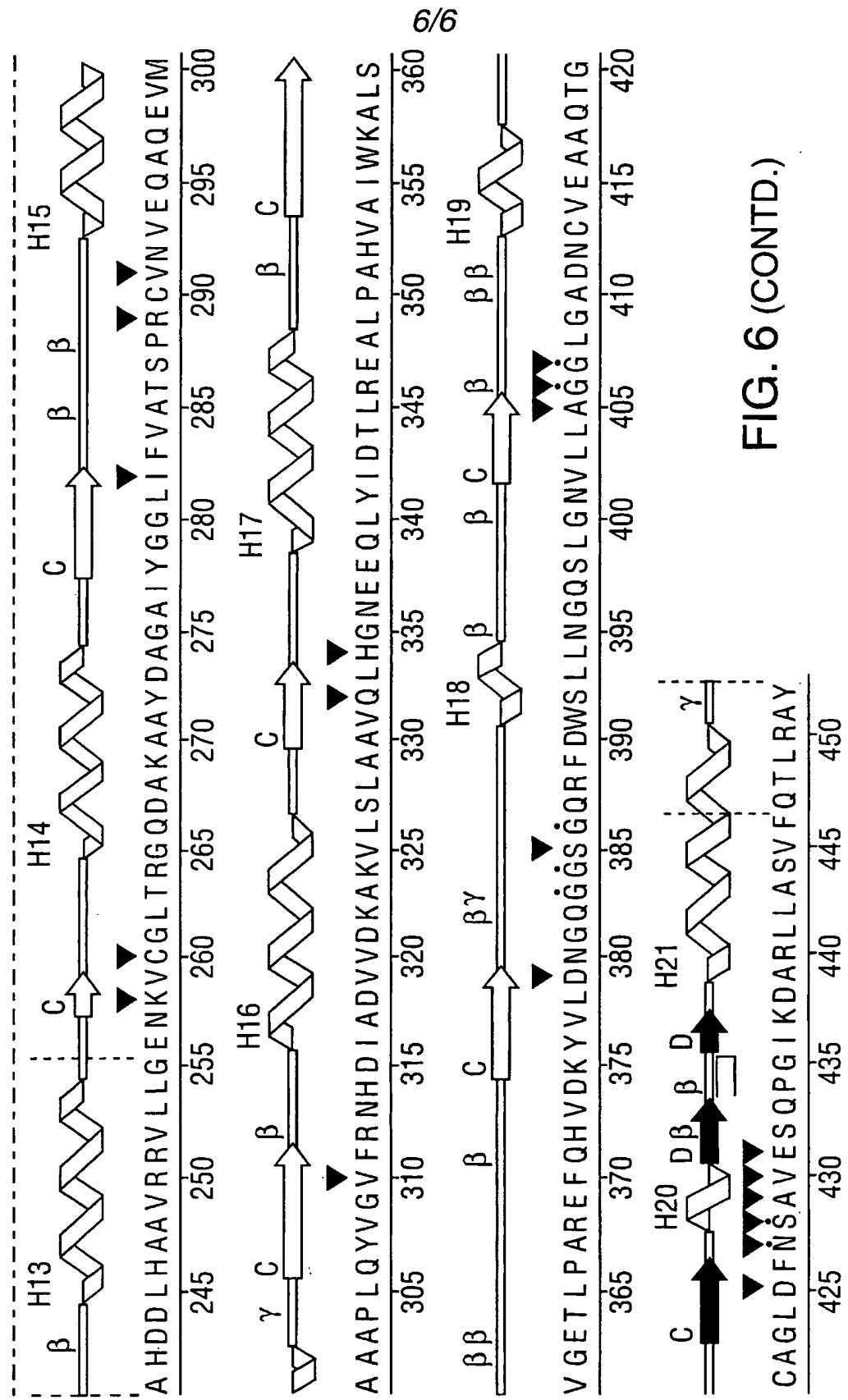


FIG. 6 (CONT'D.)